

⑫

**EUROPEAN PATENT APPLICATION**

⑪ Application number: 88400790.7

⑤ Int. Cl.<sup>4</sup>: **C 12 N 15/00**  
**A 61 K 39/29**

⑫ Date of filing: 31.03.88

⑬ Priority: 31.03.87 JP 78313/87  
04.06.87 JP 140586/87  
10.11.87 JP 283990/87

⑭ Date of publication of application:  
30.11.88 Bulletin 88/48

⑮ Designated Contracting States:  
CH DE FR GB IT LI NL SE

⑯ Applicant: MITSUBISHI CHEMICAL INDUSTRIES  
LIMITED  
5-2, Marunouchi 2-chome Chiyoda-ku  
Tokyo 100 (JP)

⑰ Inventor: Takahashi, Kazuhiro 303 Corpo Omodaka  
29 Ichijoji-Somedono-cho Sakyo-ku  
Kyoto City Kyoto Prefecture (JP)

Shibui, Tatsurou  
M806-11 San White 6-15-4 Minami-Naruse  
Machida City Tokyo (JP)

Kamizono, Michiru  
F202-24 San White 2-8-11 Tsuruma  
Machida City Tokyo (JP)

Matsui, Rie  
A211-23 San White 1-26-31 Ogawa  
Machida City Tokyo (JP)

Teranishi, Yutaka  
1-14-2 Takane  
Sagamihara City Kanagawa Prefecture (JP)

Nakanishi, Shigetada  
517-116 Iwakura-Nagatani-cho Sakyo-ku  
Kyoto City Kyoto Prefecture (JP)

Kitamura, Naomi  
2-272 Yakumo-Higashi-machi  
Moriguchi City Osaka Prefecture (JP)

⑲ Representative: Gutmann, Ernest et al  
S.C. Ernest Gutmann - Yves Plasseraud 67, boulevard  
Haussmann  
F-75008 Paris (FR)

The title of the invention has been amended (Guidelines for Examination in the EPO, A-III, 7.3).  
The application is published incomplete as filed (Article 93 (2) EPC). The point in the description or the claim(s) at which the omission obviously occurs has been left blank.

⑳ DNA molecules encoding non-A, non-B hepatitis antigens, and their use in producing said antigens.

㉑ Provided herein are a DNA fragment which contains a base sequence coding for a non-A non-B hepatitis-specific antigenic protein occurring in cells of the liver affected with non-A non-B hepatitis, an expression vector in which said DNA fragment is inserted into a cloning site present downstream from a promoter thereof, a transformant obtained by introducing said expression vector into a host, and a process for producing said antigenic protein which comprises providing said expression vector, transforming a host with said expression vector, culturing the transformed host and collecting the protein produced therein.

**EP 0 293 274 A1**

**Description****DNA FRAGMENTS CODING FOR ANTIGENS SPECIFIC TO NON-A NON-B HEPATITIS, EXPRESSION VECTORS CONTAINING SAID DNA FRAGMENTS, TRANSFORMANTS AND PROCESS FOR PRODUCING SAID ANTIGENS**

5

**BACKGROUND OF THE INVENTION****Field of the Invention:**

10 The present invention generally relates to the production of an antigen specific to non-A non-B hepatitis by recombinant DNA technology. More particularly, it relates to a DNA fragment coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis, an expression vector containing such a DNA fragment, a host transformed with such an expression vector, as well as a process for producing said antigen specific to non-A non-B hepatitis by culturing such a transformant.

15 **Description of the Prior Art:**

Among viral hepatitises, the viral entities of hepatitis type A and type B have been found and, accordingly, it has now become possible to diagnose such diseases by immunological methods.

20 Still another type of hepatitis different from the types A and B, which is called non-A non-B type hepatitis, is said to be over 90% of post-transfusion hepatitis: refer to NIPPON RINSHO (Japan Clinic), 35, 2724 (1977); J. Biol. Med., 49, 243 (1976). The pathogenic virus of the non-A non-B type hepatitis, however, has not yet been identified. Only one fact which has already been established is potential infection of human hepatitis type non-A non-B virus to chimpanzee: refer to Lancet I, 459 (1978); *ibid.*, 463 (1978).

25 Many workers have done various investigations for searching an antigen-antibody system related to the non-A non-B hepatitis by using mainly sera from patients affected with the disease; nevertheless, no definite system has been found. Under these circumstances, the diagnosis of non-A non-B hepatitis should inevitably be effected by so-called exclusion diagnosis: that is, whether or not the hepatitis of a patient is type A or type B or other hepatitis due to a virus known to cause hepatopathy, for example, CMV, HSV, EBV, etc., is first determined; and if not, the patient's hepatitis is diagnosed as non-A non-B type. Thus, such a diagnosis of non-A non-B hepatitis will require much time and labor.

30 An antigenic protein specific to non-A non-B hepatitis and useful for the direct diagnoses of the hepatitis has been purified from human and chimpanzee hepatocytes affected with non-A non-B hepatitis, and a monoclonal antibody specific to the antigen and useful for the treatment of the non-A non-B hepatitis has also been proposed: to Japanese Patent Application Laying-open (KOKAI) Nos. 176856/86 and 56196/86.

35 A large amount of such an antigenic protein specific to non-A non-B hepatitis should be required when such a protein is to be employed, for example, as a diagnostic agent. However, it is not always appropriate to purify such a large amount of the antigenic protein from chimpanzee hepatocytes affected with non-A non-B hepatitis.

40 On the other hand, in order to detect a gene coding for a specific antigen of non-A non-B hepatitis by nucleic acid hybridization and, further, to produce such an antigen specific to non-A non-B hepatitis by the recombinant DNA technology, it is essential to obtain a gene fragment coding for the antigenic protein specific to the non-A non-B hepatitis.

**SUMMARY OF THE INVENTION**

45 The present inventors have made great efforts to produce such a specific antigenic protein in a large amount by genetic engineering techniques, and finally isolated a gene fragment coding for the antigenic protein specific to non-A non-B hepatitis, said gene fragment being useful for the production of such antigens. Further, the inventors have successfully constructed an expression vector containing said gene fragment. Thus, the present invention has now been attained.

50 It is an object of the invention to provide a DNA fragment which contains a base sequence coding for an antigen specifically occurring in a host cell affected with non-A non-B hepatitis or an antigenic protein specific to non-A non-B hepatitis having physiological activities equivalent to those of said specifically occurring antigen.

Another object of the invention is to provide an expression vector having said DNA fragment introduced therinto at a cloning site downstream from a promoter of the vector.

55 A still another object of the invention is to provide a transformant obtained by transforming a host cell with said expression vector.

A further object of the invention is to provide a process for producing such an antigen specific to non-A non-B hepatitis by culturing said transformant.

60 **BRIEF DESCRIPTION OF THE DRAWINGS**

Other objects and advantages of the present invention will be apparent from the following detailed description with reference to the attached drawings, in which:

Figs. 1a-1e show the base sequence coding for an antigenic protein specific to non-A non-B hepatitis;

Fig. 2 shows the base sequence of a hybrid promoter Pac;  
 Figs. 3a-3c show the base sequence of a cDNA fragment obtained in Example 1 described hereinbelow, together with deduced amino acid sequence;  
 Figs. 4a-4c show the base sequence of cDNA containing the full length gene sequence of an antigenic protein specific to non-A non-B hepatitis, which cDNA was obtained in Example 2 described hereinbelow, the base sequence 57-1388 thereof coding for the antigenic protein specific to non-A non-B hepatitis;  
 Fig. 5 schematically illustrates the construction of a plasmid pCV44H;  
 Fig. 6 schematically illustrates the construction of a plasmid pCV44B; and  
 Fig. 7 schematically illustrates the construction of a plasmid pCZ44.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention will be described in detail hereinbelow.

According to one aspect of the invention, a DNA fragment is provided which contains a base sequence coding for an antigenic protein occurring specifically in hepatocytes affected with non-A non-B hepatitis.

Such a DNA fragment of the invention may be prepared in the following manner.

First, a liver tissue specimen derived from a human or chimpanzee individual affected with non-A non-B hepatitis is homogenized in an aqueous solution of guanidinium thiocyanate and then subjected to cesium chloride equilibrium density gradient centrifugation according to Chirgwin et al. method (Biochemistry, 18, 5294-5299 (1979)) to separate total RNA as a precipitate. After separation, the total RNA is purified by phenol extraction and ethanol precipitation.

"Individuals affected with non-A non-B hepatitis" used as sources of liver tissue specimens in the invention may include those affected with so-called type D hepatitis, which has recently been named.

It is known that mRNA of an antigen gene generally has a poly-A chain. Thus, the total RNA is subjected to oligo(dT) cellulose column chromatography in a conventional manner and poly(A)-containing RNA (poly A<sup>+</sup> RNA) is isolated as mRNA material.

A cDNA library corresponding to the poly A<sup>+</sup> RNA is then obtained from the mRNA material according to the random primer method (Y. Ebina et al., Cell, 40, 747-758 (1980)): Thus, a number of DNAs complementary to the mRNA material are randomly synthesized using any primer of e.g. about 6 bases and a reverse transcriptase.

The cDNA is methylated with a DNA methylase, e.g. EcoRI methylase, to protect cleavage sites present in the cDNA capable of being cleaved by a corresponding restriction enzyme, e.g. EcoRI. A DNA linker containing the corresponding restriction enzyme cleavage sites at both ends, e.g. EcoRI linker (CGAATTCG), is added to the methylated cDNA and, then, this cDNA is digested with the restriction enzyme, e.g. EcoRI.

The digested cDNA is then cloned into a cloning vector such as a plasmid or a  $\lambda$  phage. For example, the cDNA may be introduced into EcoRI site of  $\lambda$ gt 11 DNA, which is an expression cloning vector: refer to R.A. Young et al., Proc. Natl. Acad. Sci. U.S.A., 80, 1194-1198 (1983). The cDNA will be inserted into the  $\beta$ -gal gene on the  $\lambda$ gt 11 phage. Thus, expression of the cDNA can be easily verified by the production of a fused protein with  $\beta$ -galactosidase due to induction of the expression by the lactose operon promoter of said phage when *E. coli* transfected with said phage is cultured in a medium containing IPTG (Isopropylthio- $\beta$ -D-galactopyranoside).

The  $\lambda$ gt 11 phage incorporating the cDNA is then introduced into *E. coli* by Tomizawa et al. method in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). The thus transfected microorganism is cultured in an IPTG-containing medium.

The thus formed plaques can be easily selected by an immunological screening method using a monoclonal antibody specifically directed to non-A non-B hepatitis to obtain a desired cDNA. Such a monoclonal antibody which can be used in the immunological screening method may be prepared according to the methods described in Japanese Patent Application Laying-open Nos. 176856/86 and 56196/86. The screening methods used may include the western blotting technique described in these applications.

The plaques positive in the immunological screening test are selected to proliferate the phage by Tomizawa et al. method. DNA is purified from the grown phage by T. Maniatis et al. method in "Molecular Cloning", Cold Spring Harbor Laboratory, pp. 85 et seq. (1982), and cleaved with a suitable restriction enzyme such as EcoRI. The thus purified and digested DNA fragments can be used to determine the base sequence of a desired cDNA segment according to Maxam and Gilbert method in Methods in Enzymology, 65, 499-560 (1980); or alternatively, after further cloning the DNA fragments into M13 phage, the base sequence of such a desired cDNA segment can be determined according to the dideoxy method: Sanger et al., Proc. Natl. Acad. Sci. U.S.A., 74, 5463 (1977).

Thus, a cDNA fragment coding for an antigen specific to non-A non-B hepatitis can be obtained. However, such a DNA fragment may usually be only a portion of the gene coding for the non-A non-B hepatitis-specific antigen.

A full length cDNA coding for such a non-A non-B hepatitis-specific antigen may be obtained in the following manner.

Poly A<sup>+</sup>-mRNA is isolated and purified in a manner similar to that described above. From the poly A<sup>+</sup>-mRNA a cDNA library is obtained according to Okayama-Berg vector-primer method: Molecular and Cellular Biology, 2, 161-170 (1982).

A plasmid containing such a cDNA thus prepared is used to transform *E. coli* by any conventional method, for instance, the method D. Hanahan: J. Mol. Biol., 166, 557 (1983). The transformant ampicillin-resistant

strains are collected and screened by the colony hybridization method using the aforementioned DNA fragment as a probe. Such a probe may preferably be prepared by either the strepto-avidin method, or the nick translation method using photobiotinnucleic acids and  $^{32}\text{P}$ -nucleic acids.

The thus selected colonies containing a cDNA clone are cultured. Plasmid DNA is obtained from the cultured colony according to Birnboim *et al.* method (Nucleic Acid Res., 7, 1513 (1979)) and digested with a suitable restriction enzyme. The base sequence of a desired full-length cDNA segment is then determined according to the aforementioned Maxam and Gilbert method or, alternatively, after further cloning the digested DNA into M13 phage or pVC12 plasmid, such a base sequence is determined according to the above described Sanger *et al.* dideoxy method.

The base sequence of the full length DNA coding for an antigen specific to non-A non-B hepatitis is shown in Fig. 1, in which the symbol "-" just under the base sequence represents a corresponding base complementary to the respective base described just above each of the symbols.

Of course, DNA fragments which can be employed in the invention do not necessarily contain the same base sequence as shown in Fig. 1, but those DNA fragments in which a part of said base sequence shown in Fig. 1 has been substituted by at least one different base or deleted therefrom and those DNA fragments in which one or more additional bases have been added to the base sequence of Fig. 1 may also be included herein provided that such different DNA fragments may code for substances having physiological activities equivalent to those of the non-A non-B hepatitis-specific antigens encoded by the base sequence of Fig. 1.

According to another aspect of the invention, an expression vector is provided in which the aforementioned DNA fragment of the invention is inserted into a cloning site downstream from a promoter of this vector.

The expression vector of the invention contains a promoter in a position capable of controlling the transcription of a DNA fragment coding for a non-A non-B hepatitis-specific antigen obtained by the aforementioned method. The promoters used in the invention may be any promoter capable of expressing the DNA fragment in a host, and preferably of controlling the transcription of the fragment.

When a host used is a microorganism such as *Escherichia coli*, *Bacillus subtilis*, etc., the expression vector of the invention may preferably comprise a promoter, a ribosome binding sequence, a gene for a non-A non-B hepatitis-specific antigen, a transcription termination factor, and a gene controlling the promoter.

The promoter used may include those derived from *E. coli*, phage, etc., for example, tryptophan synthase operon (*trp*), lactose operon (*lac*), lipoprotein (*lpp*), *recA*, lambda phage *P<sub>L</sub>*, *P<sub>R</sub>*, T5 early gene *P<sub>25</sub>*, *P<sub>28</sub>* promoter, which may also be prepared by chemical synthesis. Also included herein are hybrid promoters such as *tac* (*trp:lac*), *trc* (*trp:lac*) and *Pac* (phage:*E. coli*) shown in Fig. 2.

The ribosome binding sequence may be derived from *E. coli*, phage, etc., but preferably may be those synthetically prepared, for example, those containing a consensus sequence such as

AGGAGGTTTAA.

SD sequence

The gene for a non-A non-B hepatitis-specific antigen may be directly employed without any modification. Preferably, an unnecessary base sequence (non-coding region) may be deleted by site-directed mutagenesis: BIO TECHNOLOGY, July, 636-639 (1984).

A transcription termination factor may not always be required in the expression vector of the invention. Preferably, the instant vector may contain a p-independent terminator, for example, *lpp* terminator, *trp* operon terminator, ribosomal RNA gene terminator, etc.

The expression vector may be derived from any conventional plasmid. Preferably, it may be derived from such a plasmid as replicating itself in *E. coli* or *Bacillus subtilis*, for example, pBR322- or pUB110-derived plasmid.

Desirably, these factors required for expression are arranged in the expression plasmid in the order of the promoter, the SD sequence, the structural gene of a non-A non-B hepatitis-specific antigen, and the transcription termination factor from 5' to 3'. A repressor gene required to control the transcription, a marker gene such as drug-resistant gene, and a plasmid replication origin may be arranged in any order in the expression vector.

The expression vector of the invention may be introduced into a host by any conventional method for transformation of *E. coli*, e.g., one described in Molecular Cloning, 250-253 (1982), or of *Bacillus subtilis*, e.g., one described in Molec. Gen. Genet., 168, 111-115 (1979) or Proc. Nat. Acad. Sci. U.S.A., 44, 1072-1078 (1958).

The resulting transformant may be cultured in any conventional medium, e.g. one described in Molecular Cloning, 68-73, (1972), at a temperature in the range of 28 to 42°C in both cases of *E. coli* and *Bacillus subtilis*. Preferably, it may be cultured at a temperature in the range of 28 to 30°C where no expression of heat shock proteins may be induced.

The desired protein thus produced may be easily purified from the host in conventional procedures. For example, the host cell may be crushed by lysozyme-surfactant or ultra-sonication, and the insoluble fractions which contain the desired non-A non-B hepatitis-specific antigen may be then collected by centrifugation,

solubilized in a surfactant such as 0.01% SDS, and subjected to column chromatography using a monoclonal antibody (Japanese Patent Application Laying-open (KOKAI) Nos. 56196/86 and 176856/86.

When an eukaryotic cell such as an animal cell is employed as a host, the expression vector of the invention is preferably as follows:

The promoters used in the vector of the invention for the expression in eukaryotic cells may herein include SV40 early and late promoters; promoters of apolipoprotein E and A-I genes; promoter of heat shock protein gene (Proc. Natl. Acad. Sci. U.S.A., 78, 7038-7042 (1981)); promoter of metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6511-6515 (1980)); HSV TK promoter; adenovirus promoter, such as Ad2 major late promoter (Ad2 MLP); LTR (long terminal repeat) of retrovirus; etc. SV40 promoter and promoter of metallothionein gene are preferred.

The expression vector of the invention may contain a splice sequence comprising 5' splice junction donor site, an intron and 3' splice junction acceptor site. A common base sequence is found at all the splice junction sites (exonintron junction sites); so-called GT/AG rule that any intron region always starts from two bases GT at the donor site and terminates at two bases AG of the acceptor site has been established.

The expression vector of the invention may contain one or more splice sequences as mentioned just above. Such splice sequences may be positioned upstream or downstream of the structural gene for a non-A non-B hepatitis-specific antigen.

Illustrative examples of such splice sequences may include those DNA sequences found in exons 2 and 3 of rabbit  $\beta$ -globin gene (Science, 206, 339 (1979)) and mouse metallothionein-I gene containing the promoter, exons 1, 2 and 3 and introns A and B of metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6513 (1980)). The 5' and 3' splice sites may be derived from the same or different gene; for example, a sequence in which 5' splice site contained in adenovirus DNA is linked to 3' splice site derived from the gene of Ig variable region can be employed.

The expression vector of the invention also contains a polyadenylation site downstream from the structural gene of a non-A non-B hepatitis-specific antigen. Illustrative examples of the polyadenylation sites may include those derived from SV40 DNA,  $\beta$ -globin gene or metallothionein gene. A combined site of the polyadenylation sites of  $\beta$ -globin gene and SV40 DNA may be employed in the invention.

The expression vector of the invention may also contain a dominant selective marker permitting the selection of transformants. Selective markers which can be used herein may include DHFR gene imparting MTX (methotrexate) resistance to a host; tk gene of herpes simplex virus (HSV) which permits selection of tk<sup>-</sup> strains transformed therewith in HAT medium; the gene for aminoglycoside 3'-phosphotransferase from *E. coli* transposon Tn5, which imparts to a host the resistance against 3'-deoxystreptomycin antibiotic G418; bovine papilloma virus gene permitting morphological discrimination by piled up growth; and *aprt* gene.

Alternatively, animal cells transformed with the expression vector of the invention may be selected by the cotransformation even though no selective marker is present in the vector. For this purpose, an animal cell is cotransformed with both the expression vector and a plasmid or other DNA containing a gene for such a selective marker and selected by a phenotypic trait of the gene.

Advantageously, the expression vectors may also contain a plasmid fragment having an origin of replication derived from a bacterium such as *E. coli*, since such vectors can be cloned in bacteria. Such plasmids may include pBR322, pBR327, pML, etc.

Illustrative examples of plasmid vectors used as sources of the expression vectors according to the invention may include pKCR (Proc. Natl. Acad. Sci. U.S.A., 78, 1528 (1981)), which contains SV40 early promoter, the splice sequence and polyadenylation site derived from rabbit  $\beta$ -globin gene, the polyadenylation site from SV40 early region, and the origin of replication and ampicillin resistant gene from pBR322; pKCR H2 (Nature, 307, 605 (1984)), in which the pBR322 portion of pKCR has been substituted by pBR327 fragment and the *EcoRI* site present in the exon 3 of rabbit  $\beta$ -globin gene has been converted into *HindIII* site; and pBPV MT1 containing BPV gene and metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 80, 398 (1983)).

Animal cells transformed with the expression vector of the invention may include CHO cells, COS cells, and mouse L cells, C127 cells and FM3A cells.

The introduction of the expression vector of the invention into an animal cell may be carried out by transfection, microinjection, etc. Most often, the transfection may employ  $\text{CaPO}_4$ : Virology, 52, 456-467 (1973).

Animal cells transformed by introducing the expression vector of the invention may be cultured in a suspension or solid medium by conventional methods. The culture medium used is most often MEM, RPMI1640, etc.

Proteins produced in the transformed animal cells can be separated and purified in the almost same manner as in the case of microorganisms aforementioned.

As stated, the invention provides a transformant cell obtained by introducing the expression vector of the invention into a host cell.

Also provided according to the invention is a process for producing a non-A non-B hepatitis-specific antigen comprising culturing said transformant and collecting the produced and accumulated antigen.

As stated previously, a large amount of an antigenic protein specific to non-A non-B hepatitis is required when such a protein is to be utilized as a direct diagnostic agent. According to the present invention, such an antigenic protein can be produced with a low cost and a large scale without use of infected chimpanzee hepatocytes. Prior to the present invention, it has been difficult to obtain such a large amount of a non-A non-B hepatitis-specific antigenic protein from hepatocytes of chimpanzees affected with non-A non-B hepatitis.

Further, the DNA fragment coding for an antigenic protein of non-A non-B hepatitis virus according to the present invention will be useful as a probe for detecting the gene of said antigenic protein by nucleic acid hybridization.

## 5 EXAMPLES

The following examples will be given by way of illustration but these examples in no way limit the scope of the invention without departing the concept thereof.

### EXAMPLE 1: Preparation of cDNA Fragment Coding for Antigenic Protein Specific to Non-A Non-B Hepatitis

10 Poly(A)-containing RNA was prepared from chimpanzee liver according to the guanidine thiocyanate-lithium chloride method: Cathala et al., DNA, 2, 329 (1983).

The infected liver (5 g) was taken out from a chimpanzee affected with non-A non-B hepatitis and immediately frozen by liquid nitrogen. The frozen liver was added into a Waring blender together with liquid nitrogen and ground at 3,000 rpm for 2 minutes. The ground liver specimen was further ground by a Teflon homogenizer at 5 rpm in 100 ml of a solution: 5 M guanidine thiocyanate, 10 mM EDTA, 50 mM Tris-HCl (pH 7), 8% (v/v)  $\beta$ -mercaptoethanol. The thus solubilized material (20 ml) was slowly placed on 5.7 M CsCl solution (10 ml) contained in a centrifuge tube and centrifuged at 27,000 rpm for 20 hours in Hitachi RPS 28-2 rotor. The thus precipitated RNA was collected and dissolved in 10 ml of a solution: 0.1% sodium laurylsulfate, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The RNA was extracted with phenol-chloroform and re covered by ethanol precipitation.

The thus obtained RNA (about 3.95 mg) was dissolved in 1 ml of a solution: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. The solution was incubated at 65°C for 5 minutes, and 5 M NaCl (0.1 ml) was added. The resulting mixture was subjected to chromatography on an oligo(dT) cellulose column (column volume of 0.5 ml, P-L Biochemical). The thus adsorbed poly(A)-containing mRNA was eluted with a solution: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. There was obtained about 100  $\mu$ g of poly(A)-containing mRNA.

The thus obtained poly(A)<sup>+</sup> mRNA (10  $\mu$ g) was dissolved in 50  $\mu$ l of RT buffer: 20 mM Tris-HCl (pH 8.8), 0.1 M KCl, 12 mM MgCl<sub>2</sub>, 2 mM MnCl<sub>2</sub>. To this solution, there was added 8  $\mu$ g of random primer d(N)<sub>8</sub> (P-L Biochemical). The resulting mixture was heated at 95°C for 3 minutes to denature the materials, which was then cooled gradually to room temperature to anneal the random primer with the mRNA. To the annealed mixture, there were added 10 mM 4NTP (10  $\mu$ l) and reverse transcriptase (225 units) from TAKARA SHUZO (Japan), and then water was added so as to make the total volume of the mixture to 100  $\mu$ l. Reaction was allowed to proceed at 42°C for one hour.

To the reaction mixture (50  $\mu$ l), there were added 10 mM NAD (2  $\mu$ l), 10 mM 4dNTP (10  $\mu$ l), RNase H (5 units), E. coli ligase (1 unit), E. coli DNA polymerase I (6.3 units), and 10  $\times$  T4 DNA ligase buffer (10  $\mu$ l; 0.1 M Tris-HCl, pH 7.5, 0.1 M DTT, 60 mM MgCl<sub>2</sub>) to make the total volume to 100  $\mu$ l. The mixture was allowed to react at 37°C for one hour to synthesize a double stranded DNA.

The thus obtained double stranded DNA was extracted with an equal volume of water-saturated phenol. Phenol in the aqueous layer was removed with the aid of ether followed by ethanol precipitation. The precipitate thus obtained was dissolved in 50  $\mu$ l of water, and 10  $\times$  T4 DNA polymerase buffer (10  $\mu$ l; 0.33 M Tris-acetic acid, pH 7.9, 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT), 10 mM 4dNTP (10  $\mu$ l), and T4 DNA polymerase (6 units) were added to make the total volume to 100  $\mu$ l. The mixture was reacted at 37°C for one hour. There was obtained a double stranded DNA having blunt ends, which was then extracted with phenol to remove proteins and purified by ethanol precipitation as described above. The thus purified DNA was then air dried.

45 To the purified DNA, there were added 50 mM Tris-HCl (pH 7.5), 1 mM Na<sub>2</sub>EDTA, 5 mM DTT (20  $\mu$ l), 100  $\mu$ M S-adenosyl-L-methionine (2  $\mu$ l), and 1.8 mg/ml EcoRI methylase (0.2  $\mu$ l). Reaction was effected at 37°C for 15 minutes, whereby methylating the EcoRI restriction enzyme cleavage site on the DNA fragment. The reaction mixture was then heated at 70°C for 15 minutes to deactivate the enzyme.

To the reaction mixture, there was added 3'-phosphorylated EcoRI linker (GGAATTCC) in an amount of 100 molecules thereof per molecule of the synthetic DNA. There were further added 10  $\times$  T4 DNA ligase buffer (5  $\mu$ l; 0.5 M Tris-HCl, pH 7.5, 60 mM MgCl<sub>2</sub>, 10 mM DTT), 0.1 M ATP (5  $\mu$ l), and T4 DNA ligase (5 units) to make the total volume to 50  $\mu$ l. The resulting reaction mixture was reacted at 4°C for 16 hours followed by heating at 70°C for 10 minutes to deactivate the enzyme. Then, 10  $\times$  EcoRI buffer (10  $\mu$ l; 15 M Tris-HCl, pH 7.5, 0.5 M NaCl, 60 mM MgCl<sub>2</sub>), and EcoRI (100 units) were added to make the total volume to 100  $\mu$ l, and the reaction mixture was reacted at 37°C for 2 hours to cut the linker. The reaction mixture was passed through Bio Gel A-50 (0.2 cm  $\times$  32 cm, Bio RAD). Elution was effected by a buffer: 10 mM Tris-HCl (pH, 7.5), 6 mM MgCl<sub>2</sub>. Excess EcoRI linker was removed and, thus, a double stranded cDNA having EcoRI sites at both ends thereof was purified.

To the thus obtained double stranded cDNA fragment having EcoRI sites at both ends, there were added gt 11 DNA (10  $\mu$ g) cleaved with EcoRI, 10  $\times$  T4 DNA ligase buffer (10  $\mu$ l) as described above, 0.1 M ATP (10  $\mu$ l), and T4 DNA ligase (10 units) to make the total volume to 100  $\mu$ l. The mixture was reacted at 4°C for 16 hours. Thus, said double stranded cDNA fragment was inserted into  $\lambda$ gt 11 DNA.

The  $\lambda$  phage packaging kit (PROMEGA, Biotech) was used to introduce said DNA into  $\lambda$  phage particle. The procedures for packaging were effected according to the instructions of the kit.

65 The  $\lambda$ gt 11 phage having said DNA packaged thereinto was used to transfect E. coli strain Y1090 to form

plaques according to the conventional Tomizawa et al. methods described in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). Among about 200,000 plaques, one positive clone was selected by immunological screening as described hereinbelow. A monoclonal antibody used in the immunological screening was prepared by the method described in Japanese Patent Application Laying-open (KOKAI) No. 176856/86.

*E. coli* Y1090 (R.A. Young et al., *Proc. Natl. Acad. Sci. U.S.A.*, 80, 1194-1198 (1983), which had been transfected with  $\lambda$ gt 11, was inoculated in a petri dish together with soft agar held at 42°C. The transfected cell was allowed to stand at 42°C for 5 hours. A nitrocellulose filter (S & S, BA-83, pore size of 0.2  $\mu$ m) containing 10 mM IPTG was placed on the cell in the dish and incubation was effected at 37°C for 3-4 hours. This nitrocellulose filter was lightly rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed in the TBS buffer (400 ml) containing 3% gelatine and shaken at 40°C for one hour. Thus, the nitrocellulose filter was blocked. Then, a monoclonal antibody ( $OD_{280}=4.3$ ) directed to a non-A non-B hepatitis-specific antigen was added to TBS buffer containing 1% gelatine with a dilution of 1/400. This mixture was put into a vinyl bag together with the filter in a proportion of 2 ml of the mixture per filter, and reaction was allowed to proceed at room temperature for 16 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes. A labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad) was added to TBS buffer containing 1% gelatine with a dilution of 1/1,000. This mixture and the filter were put into a vinyl bag with a proportion of 2 ml of the mixture per filter. Reaction was allowed to proceed at room temperature for 2 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes, in the same manner as described above. Color development was effected by dipping the filter and 4-chloro-1-naphthol (12 mg, Bio Rad) into 20 ml of TBS buffer containing hydrogen peroxide. After completion of the color development, the filter was thoroughly washed with water and put into a vinyl bag containing water. The bag was stored in a dark and cold place.

Thus, one positive plaque was obtained. The plaque was three times subjected to single plaque isolation. In each time, immunological screening was effected in the same manner as described above, verifying that the plaque was in fact positive.

The phage was then cultured in a large scale to purify the DNA in the following manner: First, *E. coli* Y1090 was cultured overnight in 10 ml of NZ medium prepared by adding NZ amine (10 g), NaCl (5 g) and 5 mM  $MgCl_2$  to one liter of water followed by adjusting the pH to 7.2. The culture (1 ml) was transfected with the phage, with the m.o.i. (multiplicity of infection) being 0.1. The transfected culture was allowed to stand at 37°C for 10 minutes and then transferred to one liter of NZ medium. Shaking culture was effected at 37°C for 7-8 hours until the cells were lysed. Chloroform (5 ml) was added to the culture and shaking was continued for additional 30 minutes. The culture was subjected to centrifugation at 6,500 rpm for 10 minutes to remove cell debris.

NaCl (29 g) and polyethylene glycol (70 g) were added to and thoroughly dissolved in the obtained supernatant, and the solution was allowed to stand at 4°C overnight. The precipitate was collected by centrifugation at 6,500 rpm for 20 minutes, drained thoroughly, and dissolved in 20 ml of TM buffer: 10 mM Tris-HCl (pH 7.5), 5 mM  $MgCl_2$  DNase I and RNase A were added to the solution, both with a concentration of 10  $\mu$ g/ml, and the reaction was effected at 37°C for one hour. Chloroform (20 ml) was then added to the reaction mixture and stirred; thus, polyethylene glycol was distributed in the chloroform layer which was then separated from the aqueous layer. This aqueous layer was ultra-centrifuged at 28,000 rpm for 60 minutes. Thus, a pellet of phage particles was obtained.

This pellet was dissolved in TM buffer (1 ml) and subjected to CsCl density gradient centrifugation at 33,000 rpm for 20 hours. The resultant fraction containing the phage particles ( $\rho=1.45-1.50$ ) was dialyzed overnight against TM buffer. Proteinase K was added to the dialyate in an amount of 100  $\mu$ g/ml and reaction was effected at 37°C for one hour. Thereafter, an equal volume of water-saturated phenol was added and phenol-extraction was gently effected. After centrifugation at 6,500 rpm for 10 minutes, the aqueous layer was removed, put into a dialysis tube, and dialyzed overnight against water at 4°C. Thus, about 5 mg of DNA was obtained.

Cleavage reaction of this DNA (100  $\mu$ g) with *EcoRI* (100 units) in the aforementioned buffer (100  $\mu$ l) at 37°C revealed that two cDNA segments of 390 bp and 345 bp were inserted into the phage DNA.

These two *EcoRI* fragments were re-cloned into *EcoRI* site of a cloning vector pUC 119. Base sequences of these DNA fragments were determined by the dideoxy method using commercially available primers CAGGAAACAGCTATGAC and AGTCACGACGTTGTA, respectively. The base sequence of the linking portion between these two DNA fragments was similarly determined by cutting this cDNA fragment at *BamHI* and *EcoRV* sites present therein with corresponding specific restriction enzymes, inserting the resulting *BamHI-EcoRV* DNA fragment between *BamHI* and *SmaI* sites of the plasmid pUC 119, and sequencing the fragment by the dideoxy method.

The base sequence of said cDNA fragment is shown in Fig. 3. This was a partial cDNA fragment of a gene coding for an antigenic protein specific to non-A non-B hepatitis.

#### EXAMPLE 2: Preparation of cDNA Containing the Full Length Gene Sequence

Messenger RNA was prepared as described in Example 1 and cDNA was synthesized using Okayama vector according to the conventional method described in Molecular Cloning, p. 211 et seq. The procedures used to synthesize cDNA were as follows:

To 300  $\mu$ l of a solution (10 mM Tris-HCl, pH 7.5, 6 mM  $MgCl_2$ , 10 mM NaCl), there were added 400  $\mu$ g of

pCDV 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 500 units of *Kpn*I (TAKARA SHUZO, Japan), all restriction enzymes used hereinafter having been manufactured by TAKARA SHUZO (Japan) unless otherwise noted. Reaction was effected at 37°C for 6 hours to cut the plasmid at *Kpn*I site therein. After phenol-chloroform extraction, ethanol precipitation was effected to recover DNA.

5 The DNA (about 200 µg) cleaved with *Kpn*I was added to 200 µl of a solution which was obtained by adding dTTP in a concentration of 0.25 mM to a buffer (TdT buffer): 40 mM sodium cacodylate, 30 mM Tris-HCl (pH 6.8), 1 mM CaCl<sub>2</sub>, 0.1 mM dithiothreitol (DTT). Further, 81 units of terminal deoxynucleotidyl transferase (TdT, manufactured by P-L Biochemicals) was also added. Reaction was effected at 37°C for 11 minutes. Thus, a poly(dT) chain (about 67 deoxythymidylic acid residues) was added to the 3' end at the *Kpn*I-cleaved site of pCDV 1. After phenol-chloroform extraction and ethanol-precipitation, about 100 µg of pCDV 1 DNA to which poly(dT) chain had been added was recovered from the reaction mixture.

10 The thus obtained DNA was added to 150 µl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 100 mM NaCl), and *Hpa*I (360 units) was also added, followed by reaction at 37°C for 2 hours. The reaction mixture was subjected to electrophoresis on agarose gel to separate and recover about 3.1 Kbp DNA fragment. Thus, there was obtained about 60 µg of poly(dT)-containing pCDV 1.

The thus obtained DNA was dissolved in 500 µl of a solution (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), incubated at 65°C for 5 minutes, and cooled on ice. After adding 5 M NaCl (50 µl), the mixture was subjected to chromatography on oligo(dA) cellulose column (Colaborative Research). DNA having a poly(dT) chain of sufficient length was adsorbed on the column and eluted with a solution: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Thus, there was obtained 27 µg of pCDV 1 to which poly(dT) chain had been added, abbreviated hereinafter as vector primer.

20 A linker DNA was prepared in the following manner: To 200 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 50 mM NaCl), there were added about 14 µg of pL 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 50 units of *Pst*I. Reaction was effected at 37°C for 4 hours to cut the pL 1 DNA at *Pst*I site. Phenol-chloroform extraction and ethanol precipitation of the reaction product gave about 13 µg of pL 1 DNA cleaved at *Pst*I site.

The thus obtained DNA (about 13 µg) was added to 50 µl of the TdT buffer containing dGTP at a final concentration of 0.25 mM, and 54 units of TdT (P-L Biochemicals) was also added. The mixture was incubated at 37°C for 13 minutes to add a (dG) chain (about 14 deoxyguanylic acid residues) to the 3' end at the *Pst*I-cleaved site of pL 1. After phenol-chloroform extraction, DNA was recovered by ethanol precipitation.

30 The thus obtained DNA was added to 100 µl of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 80 units of *Hind*III was also added. The mixture was incubated at 37°C for 3 hours to cut the pL 1 DNA at *Hind*III site. The reaction product was fractionated by agarose gel electrophoresis. About 0.5 Kb DNA fragment was recovered by the DEAE paper method: Dretzen et al., Anal. Biochem., 112, 295 (1981). Thus, there was obtained an oligo(dG) chain-containing linker DNA, hereinafter abbreviated simply as linker DNA.

35 The aforementioned poly(A)\* RNA (about 2 µg) prepared in the same manner as in Example 1 and the vector primer (about 1.4 µg) were dissolved in 22.3 µl of a solution: 50 mM Tris-HCl (pH 8.3), 8 mM MgCl<sub>2</sub>, 30 mM KCl, 0.3 mM DTT, 2 mM dNTP (dATP, dTTP, dGTP and dCTP) and 10 units of ribonuclease inhibitor (P-L Biochemicals). To the solution, there was added 10 units of reverse transcriptase manufactured by SEIKAGAKU KOGYO (Japan). Incubation was effected at 37°C for 40 minutes to synthesize a DNA complementary to the mRNA. After phenol-chloroform extraction and ethanol precipitation, the vector primer DNA to which a double stranded RNA-DNA had been added was recovered.

40 The thus obtained vector primer DNA containing RNA-DNA double stranded chain was dissolved in 20 µl of TdT buffer containing 60 µM dCTP and 0.2 µg poly(A). After adding 14 units of TdT (P-L Biochemical), the mixture was incubated at 37°C for 8 hours to add a (dC) chain of 12 deoxycytidylic acid residues to the 3' end of the cDNA. The reaction product was extracted with phenol-chloroform and precipitated with ethanol to recover a cDNA-vector primer DNA to which a (dC) chain had been added.

45 The thus obtained (dC) chain-containing cDNA-vector primer DNA was dissolved in 400 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 20 units of *Hind*III was also added. The mixture was incubated at 37°C for 2 hours to cut the DNA at *Hind*III site. The reaction product was extracted with phenol-chloroform and precipitated with ethanol. Thus, there was obtained 0.5 pmole of a (dC) chain-containing cDNA-vector primer DNA.

50 The thus obtained (dC) chain-containing cDNA-vector primer DNA (0.08 pmole) and the aforementioned linker DNA (0.16 pmole) were dissolved in 40 µl of a solution: 10 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 1 mM EDTA. The resulting solution was incubated at 65°C for 10 minutes, at 42°C for 25 minutes, and then at 0°C for 30 minutes. The reaction mixture was adjusted to 20 mM Tris-HCl (pH 7.5), 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M KCl and 0.1 mM β-NAD in a total volume of 400 µl.

55 To the reaction mixture, there was added 10 units of *E. coli* DNA ligase (New England Biolabs), followed by incubation overnight at 11°C. After adjusting the concentrations of dNTP and β-NAD in the reaction mixture to 40 µM and 0.15 mM, respectively, by supplementing necessary reagents, 5 units of *E. coli* DNA ligase, 7 units of *E. coli* DNA polymerase I (P-L Biochemicals) and 2 units of *E. coli* ribonuclease H (P-L Biochemicals) were added to the reaction mixture. The mixture was incubated at 12°C for one hour and then at 25°C for one hour.

60 In the course of the above reactions, a recombinant DNA containing the cDNA was cyclized and the RNA portion of the RNA-DNA double stranded chain was substituted by DNA. Thus, a desired recombinant plasmid containing a complete double-stranded DNA was produced.



The recombinant plasmid was used to transform competent cells of *E. coli* strain MC1064 prepared by conventional methods. Approximately 50,000 transformants were fixed on a nitrocellulose filter. These colonies were screened according to the colony hybridization method described in Molecular Cloning, Cold Spring Harbor Laboratory, p. 329 et seq. (1982) using the cDNA fragment obtained in Example 1 as a <sup>32</sup>P-labelled probe. Thus, three clones showed strong hybridization at 42°C.

These positive clones were analyzed in detail by Southern method: J. Mol. Biol., 98, 503 (1975). There was obtained the desired full length cDNA of a gene coding for an antigenic protein specific to non-A non-B hepatitis. The base sequence of the cDNA is shown in Fig. 4.

The expression vector containing the full length cDNA was designated as pCDVCL-I.

### EXAMPLE 3: Preparation of Expression Vector and Transformant and Expression of Specific Antigen

#### A. Preparation of Expression Vector and Transformant

##### 1) Modification of N-terminus (Fig. 5):

i) In 100 µl of a buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), pCDVCL-I (5 µg) was digested with PvuI (10 units) at 37°C for 2 hours. The reaction mixture was heated at 75°C for 15 minutes to deactivate the enzyme, dialyzed against water, and dried. The cleaved plasmid DNA was treated with T4 DNA polymerase (4 units) in 40 µl of a system: 33 mM Tris-acetic acid (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate and 0.5 mM dithiothreitol, to which 2 mM 4-deoxytriphosphate had been added; thus, the 3' protruding end of the plasmid DNA was filled in to produce a blunt end. The thus treated mixture was heated at 70°C for 10 minutes to deactivate the enzyme, dialyzed against water, and dried. The thus obtained plasmid DNA was then stored in the form of an aqueous solution (50 µl). This plasmid DNA fragment is hereinafter designated as Fragment I.

ii) On the other hand, pCDVCL-I (20 µg) was digested with NcoI and HindIII (each 20 units) at 37°C for 2 hours in 100 µl of a buffer: 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 6 mM MgCl<sub>2</sub>. The plasmid DNA was subjected to 5% acrylamide gel electrophoresis at 10 V/cm for 1.5 hours in a buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA. The gel was stained with 0.05% aqueous ethidium bromide solution and two gel slices corresponding to DNA fragments of larger molecular weights were excised from the gel under ultraviolet radiation at 340 nm. The gel slices were crushed by means of a glass rod, suspended into 4 ml of a buffer for DNA extraction (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% sodium laurylsulfate), and allowed to stand overnight at 37°C to extract DNA from the gel. The materials were subjected to centrifugation at 10,000 rpm for 15 minutes to eliminate larger gel pieces, and passed through a glass filter to remove smaller gel pieces. The DNA was purified by effecting ethanol precipitation three times and stored in the form of an aqueous solution (200 µl). This plasmid DNA fragment is hereinafter designated as Fragment II.

iii) A primer of the DNA portion to be modified as shown below (51 bases) was synthesized by a DNA synthesizer, NIKKAKI (Japan), Applied Biosystem MODEL 380A. The synthesized DNA was overnight reacted with concentrated aqueous ammonia at 55°C to deprotect and purified by reversed HPLC before use.

Primer HindIII  
ACAACAGATCTAAGCTTATGGCAGTTACAACAAGATTAA  
(Original sequence) (-----A--A-G-----G-----TC-C--G-

CATGGTTGCATG wherein x represents a  
base substitution.  
-----)

The synthetic primer (150 pmole) was treated with T4 polynucleotide kinase (20 units) in 10 µl of a kinase buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol) to phosphorylate the 5' end thereof.

iv) Fragment I (0.05 pmole), Fragment II (0.05 pmole) and 5'-phosphorylated primer (45 pmole) were added to 12 µl of 5 × polymerase-ligase buffer (0.5 M NaCl, 32.5 mM Tris-HCl, pH 7.5, 40 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol) to make the total volume of the mixture 34.8 µl. The mixture was boiled at 100°C for 3 minutes, immediately after which it was placed in a thermostat at 30°C and allowed to stand for 30 minutes. The mixture was allowed to stand at 4°C for 30 minutes and then on ice for 10 minutes to form a heteroduplex.

To an aqueous solution (11.6 µl) containing the heteroduplex, there were added 2.5 mM 4-deoxynucleotide

triphosphate (2 µl), 10 mM ATP (2 µl), Klenow enzyme (2 units) and T4 DNA ligase (0.5 units) to form a mixture of 20 µl in total volume. The mixture was reacted overnight at 16°C to cyclize the DNA.

An aqueous solution (2 µl) containing the circular DNA was used to transform *E. coli* HB101 strain according to conventional methods. Plasmids were separated from the transformant and purified in conventional manners. The plasmid was cleaved with restriction enzyme *Hind*III and subjected to 5% acrylamide gel electrophoresis. Thus, two separate fragments were collected as desired modified, variant plasmids. Since resulting variant plasmids might often be admixed with original wild-type plasmids, the thus obtained variant plasmids were again employed to transform *E. coli* HB101 so as to purify the plasmid.

Thus, a purified plasmid pCV44H was obtained (Fig. 5).

## II) Modification of C-terminus (Fig. 6):

i) Plasmid pCDVCL-I (5 µg) was treated in the same manner as in I) i) described above to produce Fragment I.

ii) Plasmid pCDVCL-I (20 µg) was treated in the same manner as in I) ii) described above except that *Nco*I and *Nsi*I (each 5 units) were employed. Thus, Fragment II was produced.

iii) In the same manner as in I) iii) described above, the following primer (46 bases) was synthesized and the 5' end thereof was phosphorylated.

```

                                BglII  SalI
Primer          GCACAAGGAAAAAATGAGATCTGTCCACGGTTCACGTA
                                xxxxxxxxxxxxxx
(Original sequence) (-----AGATATGTGAA*A-----)

```

AATTTCC wherein x represents a base substitution and \* represents -----) an addition.

iv) The Fragment I and II and the 5'-phosphorylated primer obtained above in II) i) to iii) were treated in the same manner as in I) iv) described above. Thus, plasmid pCV44B was obtained (Fig. 6).

## III) Introduction of cDNA coding for specific antigen into expression vector (Fig. 7):

i) In 100 µl of a buffer H (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), 10 µg (about 3 pmole) of pCV44H was cut with *Hind*III (20 units) and *Sac*I (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis. Thus, a 467 bp DNA fragment coding for the N-terminus of the specific antigen was separated and purified. This fragment is hereinafter designated as Fragment N.

ii) In 100 µl of the buffer H, 10 µg (about 3 pmole) of pCV44B was cleaved with *Bgl*II (20 units) and *Sac*I (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis to isolate and purify a 836 bp DNA fragment coding for the C-terminus of the specific antigen. The thus obtained fragment is hereinafter designated as Fragment C.

iii) In 20 µl of buffer H, 2 µg (about 1 pmole) of an expression vector pUSΔH was cut with *Hind*III (2 units) and *Bgl*II (2 units) at 37°C for 2 hours. The reaction mixture was extracted with an equal volume of water-saturated phenol to remove proteins. After extracting the phenol with ether, the reaction mixture was dialyzed against water to desalt, and concentrated by a vacuum pump. Thus, there was obtained 10 µl of an aqueous solution containing an expression vector fragment HB.

iv) Fragment N (0.5 pmole), Fragment C (0.5 pmole) and the expression vector fragment HB (0.1 pmole) were mixed and reacted with T4 DNA ligase (1 unit) at 4°C for 16 hours in 10 µl of a buffer (10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 6 mM MgCl<sub>2</sub>, 1 mM ATP). The reaction mixture (3 µl) was used to transform commercially available *E. coli* JM109 competent cell according to conventional methods. The resulting transformants were selected in L broth plate (bactopeptone 10 g, yeast extract 5 g, NaCl 10 g, agar 15 g per liter) containing 20 µg/ml ampicillin. Thus, there was obtained an expression vector pCZ44 containing the specific antigen gene inserted therein (Fig. 7).

## B. Expression of Specific Antigen

*E. coli* strain JM109 possessing pCZ44 was cultured overnight at 30°C in L broth. The culture was inoculated in a fresh L broth with a dilution of 1/50 and cultured with shaking at 30°C for 2 hours. After IPTG (Isopropylthio-β-D-galactopyranoside) was added to the medium in a concentration of 2 mM, shaking culture was continued at 30°C for further 3 hours. The cells were collected by centrifugation at 6,500 rpm for 10 minutes and suspended in a buffer (0.9% NaCl, 10 mM Tris-HCl, pH 7.5) to store.

## C. Verification of Expression of Specific Antigen

The thus obtained cell culture (0.3 ml) was subjected to 10% SDS polyacrylamide gel electrophoresis at 120 V for one hour in a buffer (Tris 3g/l, glycine 14.4 g/l, 0.1% SDS). The gel was removed, placed on a nitrocellulose filter, interposed between filter papers and electrophoresed at 5 V/cm, 4°C in a buffer (Tris 3g/l, glycine 14.4 g/l) to transferred proteins in the gel onto the nitrocellulose filter. The nitrocellulose filter was rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed into 400 ml of TBS buffer containing 3% gelatine and shaken at 40°C for one hour to block the nitrocellulose filter.

To TBS buffer containing 1% gelatine, there was added a monoclonal antibody directed to a non-A non-B hepatitis-specific antigen ( $OD_{280} = 4.3$ ) with a dilution of 1/400. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 16 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

To TBS buffer containing 1% gelatine, there was added a labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad), with a dilution of 1/1000. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 2 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

Color formation was effected by immersing the filter into 20 ml of TBS buffer containing 12 mg of 4-chloro-1-naphthol (Bio Rad) and hydrogen peroxide. After completion of color formation, the filter was thoroughly washed with water, put into a vinyl bag containing water, and stored in a dark and cold place.

Such a test effected showed that a protein reacting with

The monoclonal antibody was found at the same position (44 Kd) as found in the case of the specific antigen derived from infected chimpanzee liver. This verifies that such a specific antigen can be in fact expressed in *E. coli*. The invention thus also relates to a process for the *in vitro* diagnosis of NON-A NON-B hepatitis, which comprises contacting a liver sample and/or a serum sample taken from a patient possibly infected with a NON-A NON-B hepatitis with the protein whose formula appears in claim 3 hereafter or a part thereof for a time and under conditions sufficient to allow for the production of a complex between said protein or part thereof with the antibodies contained in the patient sample and detecting the presence of the immunologic complex, particularly when the patient is suffering from NON-A NON-B hepatitis.

Any part of said protein, or any recombinant, produced by genetic engineering and including the amino acid sequence of said protein or part of said protein can be substituted for above-said protein, it being understood that the said recombinant protein or part of said protein are specifically recognized by the same antibodies as those which recognize said protein.

In other words the invention relates to all recombinant proteins or protein fragments which bind to antibodies contained in a liver extract or serum sample, or both, and originating from a patient suffering from NON-A NON-B hepatitis.

The invention also relates to a process for detecting *in vitro* an infection by a NON-A NON-B hepatitis virus, which process comprises contacting the DNA of claim 4, or a fragment thereof, under suitable hybridization conditions, with a sample of liver extract and/or serum sample originating from the patient to be diagnosed and in which the nucleic acid components had previously been made accessible to hybridization, to form a hybridization product between said DNA of claim (probe) and the viral DNA of a NON-A NON-B hepatitis B virus, and detecting said hybridization product, particularly in the case where the patient is indeed infected with a NON-A NON-B virus.

## Claims

1. A DNA fragment which contains a base sequence coding for a non-A non-B hepatitis-specific antigenic protein occurring in cells of the liver affected with non-A non-B hepatitis.
2. The DNA fragment in accordance with claim 1, in which the cells of the liver are derived from a human or chimpanzee individual.
3. The DNA fragment in accordance with claim 1, in which the antigenic protein specific to non-A non-B hepatitis has the whole or a part of the amino acid sequence represented by the formula:

0 293 274

10  
Met Ala Val Thr Thr Arg Leu Thr Trp Leu His Glu Lys Ile Leu  
20 30  
Gln Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly  
40  
Ser Val His Gly Phe His Asn Gly Val Leu Leu Asp Arg Cys Cys  
50 60  
Asn Gln Gly Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile  
70  
Ile Gly Ala Tyr Ala Glu Glu Gly Tyr Gln Glu Arg Lys Tyr Ala  
80 90  
Ser Ile Ile Leu Phe Ala Leu Gln Glu Thr Lys Ile Ser Glu Trp  
100  
Lys Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val  
110 120  
Ala Lys Tyr Asn Ser Pro Thr Asn Phe Gln Ile Asp Gly Arg Asn

0 293 274

Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu Asn Leu Gly Leu 130  
Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg 140 150  
Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu 160  
Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr Tyr Glu Pro Tyr 170 180  
Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu Leu Gly Pro Ile 190  
Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe 200 210  
Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr 220  
Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys 230 240  
Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu 250  
Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile 260 270  
Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu 280  
Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu 290 300  
Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser 310  
Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile 320 330  
Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu 340  
Thr His Val Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile 350 360  
Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val 370  
Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser 380 390

0 293 274

400  
 Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu  
 410  
 Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe  
 420  
 430  
 Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu  
 440  
 Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*.

4. The DNA fragment in accordance with claim 1, in which the base sequence comprises the whole or a part of the base sequence represented by the formula:

	10	20	30	40
5'	ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG CAT GAA AAG ATC			
3'	---	---	---	---
	50	60	70	80
	CTG CAA AAT CAT TTT GGA GGG AAG CGG CTT AGC CTT CTC TAT			
	---	---	---	---
	90	100	110	120
	AAG GGT AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG CTT GAC			
	---	---	---	---
	130	140	150	160
	AGA TGT TGT AAT CAA GGG CCT ACT CTA ACA GTG ATT TAT AGT			
	---	---	---	---
170	180	190	200	210
	GAA GAT CAT ATT ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG			
	---	---	---	---
	220	230	240	250
	GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT GCA CTT CAA GAG			
	---	---	---	---
	260	270	280	290
	ACT AAA ATT TCA GAA TGG AAA CTA GGA CTA TAT ACA CCA GAA			
	---	---	---	---
	300	310	320	330
	ACA CTG TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA ACT			
	---	---	---	---

0 293 274

340	350	360	370
AAT TTC CAG	ATA GAT GGA	AGA AAT AGA	AAA GTG ATT ATG GAC
---	---	---	---
380	390	400	410
TTA AAG ACA	ATG GAA AAT	CTT GGA CTT	GCT CAA AAT TGT ACT
---	---	---	---
430	440	450	460
ATC TCT ATT	CAG GAT TAT	GAA GTT TTT	CGA TGC GAA GAT TCA
---	---	---	---
470	480	490	500
CTG GAC GAA	AGA AAG ATA	AAA GGG GTC	ATT GAG CTC AGG AAG
---	---	---	---
510	520	530	540
AGC TTA CTG	TCT GCC TTG	AGA ACT TAT	GAA CCA TAT GGA TCC
---	---	---	---
550	560	570	580
CTG GTT CAA	CAA ATA CGA	ATT CTG CTG	CTG GGT CCA ATT GGA
---	---	---	---
590	600	610	620
GCT GGG AAG	TCT AGC TTT	TTC AAC TCA	GTG AGG TCT GTT TTC
---	---	---	---
640	650	660	670
CAA GGG CAT	GTA ACG CAT	CAG GCT TTG	GTG GGC ACT AAT ACA
---	---	---	---
680	690	700	710
ACT GGG ATA	TCT GAG AAG	TAT AGG ACA	TAC TCT ATT AGA GAC
---	---	---	---
720	730	740	750
GGG AAA GAT	GGC AAA TAC	CTG CCA TTT	ATT CTG TGT GAC TCA
---	---	---	---
760	770	780	790
CTG GGG CTG	AGT GAG AAA	GAA GGC GGC	CTG TGC ATG GAT GAC
---	---	---	---
800	810	820	830
ATA TCC TAC	ATC TTG AAC	GGT AAC ATT	CGT GAT AGA TAC CAG
---	---	---	---
850	860	870	880
TTT AAT CCC	ATG GAA TCA	ATC AAA TTA	AAT CAT CAT GAC TAC
---	---	---	---

0 293 274

	890		900		910		920						
ATT	GAT	TCC	CCA	TCG	CTG	AAG	GAC	AGA	ATT	CAT	TGT	GTG	GCA
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	930		940		950		960						
TTT	GTA	TTT	GAT	GCC	AGC	TCT	ATT	GAA	TAC	TTC	TCC	TCT	CAG
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	970		980		990		1000						
ATG	ATA	GTA	AAG	ATC	AAA	AGA	ATT	CGA	AGG	GAG	TTG	GTA	AAC
---	---	---	---	---	---	---	---	---	---	---	---	---	---
1010		1020		1030		1040		1050					
GCT	GGT	GTG	GTA	CAT	GTG	GCT	TTG	CTC	ACT	CAT	GTG	GAT	AGC
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1060		1070		1080		1090						
ATG	GAT	CTG	ATT	ACA	AAA	GGT	GAC	CTT	ATA	GAA	ATA	GAG	AGA
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1100		1110		1120		1130						
TGT	GTG	CCT	GTG	AGG	TCC	AAG	CTA	GAG	GAA	GTC	CAA	AGA	AAA
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1140		1150		1160		1170						
CTT	GGA	TTT	GCT	CTT	TCT	GAC	ATC	TCG	GTG	GTT	AGC	AAT	TAT
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1180		1190		1200		1210						
TCC	TCT	GAG	TGG	GAG	CTG	GAC	CCT	GTA	AAG	GAT	GTT	CTA	ATT
---	---	---	---	---	---	---	---	---	---	---	---	---	---
1220		1230		1240		1250		1260					
CTT	TCT	GCT	CTG	AGA	CGA	ATG	CTA	TGG	GCT	GCA	GAT	GAC	TTC
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1270		1280		1290		1300						
TTA	GAG	GAT	TTG	CCT	TTT	GAG	CAA	ATA	GGG	AAT	CTA	AGG	GAG
---	---	---	---	---	---	---	---	---	---	---	---	---	---
	1310		1320		1330								
GAA	ATT	ATC	AAC	TGT	GCA	CAA	GGA	AAA	AAA	3'			
---	---	---	---	---	---	---	---	---	---	5'			

wherein the sign "-" represents a base complementary to the base shown just above each sign.



5. An expression vector in which a DNA fragment containing a base sequence coding for a non-A non-B hepatitis-specific antigen is introduced into a cloning site present downstream from a promoter of said vector.

6. The expression vector in accordance with claim 5, in which the promoter is controllable by a regulatory factor.

7. The expression vector in accordance with claim 5, in which the promoter operates in a microorganism.

8. The expression vector in accordance with claim 5, in which the promoter operates in an eukaryote.

9. A transformant obtained by transforming a host with an expression vector in which a DNA fragment containing a base sequence coding for a non-A non-B hepatitis-specific antigen is introduced into a cloning site present downstream from a promoter of said vector.

10. The transformant in accordance with claim 9, in which the host is Escherichia coli or Bacillus subtilis.

11. A process for producing an antigen specific to non-A non-B hepatitis comprising introducing a DNA fragment containing a base sequence coding for said specific antigen into a cloning site present downstream from a promoter of a vector for expression, introducing the expression vector containing said DNA fragment into a host, culturing said transformed host, and collecting the produced and accumulated antigen.

5

10

15

20

25

30

35

40

45

50

55

60

65

0293274

Fig. 1a

```

      10      20      30
5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG
3'

      40      50      60
   CAT GAA AAG ATC CTG CAA AAT CAT TTT GGA

      70      80      90
   GGG AAG CGG CTT AGC CTT CTC TAT AAG GGT

     100     110     120
   AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG

     130     140     150
   CTT GAC AGA TGT TGT AAT CAA GGG CCT ACT

     160     170     180
   CTA ACA GTG ATT TAT AGT GAA GAT CAT ATT

     190     200     210
   ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG

     220     230     240
   GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT

     250     260     270
   GCA CTT CAA GAG ACT AAA ATT TCA GAA TGG

     280     290     300
   AAA CTA GGA CTA TAT ACA CCA GAA ACA CTG

```

Fig. 1b

310                      320                      330  
TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC

340                      350                      360  
CCA ACT AAT TTC CAG ATA GAT GGA AGA AAT

370                      380                      390  
AGA AAA GTG ATT ATG GAC TTA AAG ACA ATG

400                      410                      420  
GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT

430                      440                      450  
ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA

460                      470                      480  
TGC GAA GAT TCA CTG GAC GAA AGA AAG ATA

490                      500                      510  
AAA GGG GTC ATT GAG CTC AGG AAG AGC TTA

520                      530                      540  
CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT

550                      560                      570  
GGA TCC CTG GTT CAA CAA ATA CGA ATT CTG

580                      590                      600  
CTG CTG GGT CCA ATT GGA GCT GGG AAG TCT

0293274

Fig. 1c

	610	620	630
<u>AGC</u>	<u>TTT</u>	<u>TTC</u>	<u>AAC</u>
<u>TCA</u>	<u>GTG</u>	<u>AGG</u>	<u>TCT</u>
<u>GTT</u>	<u>TTC</u>		
	640	650	660
<u>CAA</u>	<u>GGG</u>	<u>CAT</u>	<u>GTA</u>
<u>ACG</u>	<u>CAT</u>	<u>CAG</u>	<u>GCT</u>
<u>TTG</u>	<u>GTG</u>		
	670	680	690
<u>GGC</u>	<u>ACT</u>	<u>AAT</u>	<u>ACA</u>
<u>ACT</u>	<u>GGG</u>	<u>ATA</u>	<u>TCT</u>
<u>GAG</u>	<u>AAG</u>		
	700	710	720
<u>TAT</u>	<u>AGG</u>	<u>ACA</u>	<u>TAC</u>
<u>TCT</u>	<u>ATT</u>	<u>AGA</u>	<u>GAC</u>
<u>GGG</u>	<u>AAA</u>		
	730	740	750
<u>GAT</u>	<u>GGC</u>	<u>AAA</u>	<u>TAC</u>
<u>CTG</u>	<u>CCA</u>	<u>TTT</u>	<u>ATT</u>
<u>CTG</u>	<u>TGT</u>		
	760	770	780
<u>GAC</u>	<u>TCA</u>	<u>CTG</u>	<u>GGG</u>
<u>CTG</u>	<u>AGT</u>	<u>GAG</u>	<u>AAA</u>
<u>GAA</u>	<u>GGC</u>		
	790	800	810
<u>GGC</u>	<u>CTG</u>	<u>TGC</u>	<u>ATG</u>
<u>GAT</u>	<u>GAC</u>	<u>ATA</u>	<u>TCC</u>
<u>TAC</u>	<u>ATC</u>		
	820	830	840
<u>TTG</u>	<u>AAC</u>	<u>GGT</u>	<u>AAC</u>
<u>ATT</u>	<u>CGT</u>	<u>GAT</u>	<u>AGA</u>
<u>TAC</u>	<u>CAG</u>		
	850	860	870
<u>TTT</u>	<u>AAT</u>	<u>CCC</u>	<u>ATG</u>
<u>GAA</u>	<u>TCA</u>	<u>ATC</u>	<u>AAA</u>
<u>TTA</u>	<u>AAT</u>		
	880	890	900
<u>CAT</u>	<u>CAT</u>	<u>GAC</u>	<u>TAC</u>
<u>ATT</u>	<u>GAT</u>	<u>TCC</u>	<u>CCA</u>
<u>TCG</u>	<u>CTG</u>		

0293274

Fig. 1d

	910	920	930
AAG	GAC	AGA	ATT CAT TGT GTG GCA TTT GTA
	940	950	960
TTT	GAT GCC AGC TCT	ATT GAA TAC	TTC TCC
	970	980	990
TCT	CAG ATG ATA GTA	AAG ATC AAA	AGA ATT
	1000	1010	1020
CGA	AGG GAG TTG GTA	AAC GCT GGT	GTG GTA
	1030	1040	1050
CAT	GTG GCT TTG CTC	ACT CAT	GTG GAT AGC
	1060	1070	1080
ATG	GAT CTG ATT ACA	AAA GGT GAC	CTT ATA
	1090	1100	1110
GAA	ATA GAG AGA TGT	GTG CCT	GTG AGG TCC
	1120	1130	1140
AAG	CTA GAG GAA GTC	CAA AGA AAA	CTT GGA
	1150	1160	1170
TTT	GCT CTT TCT GAC	ATC TCG	GTG GTT AGC
	1180	1190	1200
AAT	TAT TCC TCT GAG	TGG GAG CTG	GAC CCT

0293274

Fig. 2

AAAAATTTATTTGCTTTCAGGAAAATTTTCTGT  
TTTTTAAATAAACGAAAGTCCTTTTAAAAAGACA

ATAATGTGTGGAATTGTGAGCGGATAACAATTC  
TATTACACACCTTAACACTCGCCTATTGTAAAG

0293274

Fig. 1e

1210 1220 1230  
GTA AAG GAT GTT CTA ATT CTT TCT GCT CTG

1240 1250 1260  
AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC

1270 1280 1290  
TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG

1300 1310 1320  
AAT CTA AGG GAG GAA ATT ATC AAC TGT GCA

1330  
CAA GGA AAA AAA 3'  
5'

Fig. 3a

250 CTA GGA CTA TAT ACA CCA GAA ACA CTG TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA 300  
 Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val Ala Lys Tyr Asn Ser Pro  
 260 270 280 290 300  
 310 ACT AAT TTC CAG ATA GAT GGA AGA AAT AGA AAA GTG ATT ATG GAC TTA AAG ACA ATG GAA 360  
 Thr Asn Phe Gln Ile Asp Gly Arg Asn Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu  
 320 330 340 350 360  
 370 AAT CTT GGA CTT GCT CAA AAT TGT ACT ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC 420  
 Asn Leu Gly Leu Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg Cys  
 380 390 400 410 420  
 430 GAA GAT TCA CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG AGC TTA CTG 480  
 Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu Leu Arg Lys Ser Leu Leu  
 440 450 460 470 480  
 490 TCT GCC TTG AGA ACT TAT GAA CCA CCA TAT GGA TCC CTG GTT CAA CAA ATA CGA ATT CTG CTG 540  
 Ser Ala Leu Arg Thr Tyr Glu Pro Tyr Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu  
 500 510 520 530 540  
 550 CTG GGT CCA ATT GGA GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC CAA 600  
 Leu Gly Pro Ile Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe Gln  
 560 570 580 590 600



Fig. 3b

610	620	630	640	650	660
GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA ACT GAG ATA TCT GAG AAG TAT					
Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr Gly Ile Ser Glu Lys Tyr					
670	680	690	700	710	720
AGG ACA TAC TCT ATT AGA GAC GGG AAA GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT GAC					
Arg Thr Tyr Ser Ile Arg Asp Gly Lys Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp					
730	740	750	760	770	780
TCA CTG GGG CTG AGT GAG AAA GAA GGC GGC CTG TGC ATG GAT GAC ATA TCC TAC ATC TTG					
Ser Leu Gly Leu Ser Glu Lys Glu Lys Glu Gly Gly Leu Cys Met Asp Ile Ser Tyr Ile Leu					
790	800	810	820	830	840
AAC GGT AAC ATT CGT GAT AGA TAC CAG TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT CAT					
Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu Ser Ile Lys Leu Asn His					
850	860	870	880	890	900
CAT GAC TAC ATT GAT TCC CCA TCG CTG AAG GAC AGA ATT CAT TGT GTG GCA TTT GTA TTT					
His Asp Tyr Ile Asp Ser Pro Ser Leu Lys Asp Arg Ile His Cys Val Ala Phe Val Phe					
910	920	930	940	950	960
GAT GCC AGC TCT ATT GAA TAC TTC TCC TCT CAG ATG ATA GTA AAG ATC AAA AGA ATT CGA					
Asp Ala Ser Ser Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile Arg					
970	980	990	1000	1010	1020
AGG GAG TTG GTA AAC GCT GGT GTG GTA CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC ATG					
Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu Thr His Val Asp Ser Met					

Fig. 3c

1030	1040	1050	1060	1070	1080
GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC AAG					
Asp Leu Ile Thr Lys Gly Asp Leu Ile Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys					
1090	1100	1110	1120	1130	1140
CTA GAG GAA GTC CAA AGA AAA CTT GGA TTTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT					
Leu Glu Glu Val Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser Asn					
1150	1160	1170	1180	1190	1200
TAT TCC TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GAT CTA ATT CTT TCT GCT CTG AGA					
Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu Ile Leu Ser Ala Leu Arg					
1210	1220	1230	1240	1250	1260
CGA ATG CTA TGG GCT GCA GAT GAC TTC TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG AAT					
Arg Met Leu Trp Ala Ala Asp Asp Phe Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn					
1270	1280	1290	1300		
CTA AGG GAG GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA TAG					
Leu Arg Glu Glu Ile Ile Asn Cys Ala Gln Gly Lys Lys ***					

Fig. 4a

```

5' GGGGGGCTAC CCTCAGCTCT AGCTCATACT ACAGACAGTA CAACAGATCA AGAAGTATGG CAGTGACAAC TCGTTTGACA 80
3' CCCCCCGATG GGAGTCGAGA TCGAGTATGA TGTCTGTCTAT GTTGCTTAGT TCTTCATACC GTCACCTGTTG AGCAAACTGT

90 TGGTTGCATG AAAAGATCCT GCAAAATCAT TTTGGAGGGA AGCGGCTTAG CCTTCTCTAT AAGGGTAGTG TCCATGGATT 160
ACCAACGTAC TTTTCTAGGA CGTTTATGA AAACCTCCCT TCGCCGATC GGAAGAGATA TTCCCATCAC AGGTACCTAA

170 CCATAATGGA GTTTTGCTTG ACAGATGTTG TAATCAAGGG CCTACTCTAA CAGTGATTTA TAGTGAAGAT CATATTATTG
GGTATTACCT CAAAACGAAC TGTCTACAAC ATTAGTTCCC GGATGAGATT GTCACTAAAT ATCACTTCTA GTATAATAAC

250 GAGCATATGC AGAAGAGGGT TACCAGGAAA GAAAGTATGC TTCCATCATC CTTTTTGCAC TTCAAGAGAC TAAAATTTC
CTCGTATACG TCTTCTCCCA ATGGTCCTTT CTTTCATACG AAGGTAGTAG GAAAAACGTG AAGTTCTCTG ATTTTAAAGT

330 GAATGGAAC TAGGACTATA TACACCAGAA ACACGTGTTT GTTGTGACGT TGCAAAAATAT AACTCCCCAA CTAATTTC
CTTACCTTTG ATCCTGATAT ATGTGGTCTT TGTGACAAAA CAACACTGCA ACGTTTTATA TTGAGGGGTT GATTAAAGGT

410 GATAGATGGA AGAATAGAA AAGTGATTAT GGACTTAAAG ACAATGGAAA ATCTTGGACT TGCCTCAAAAT TGCTACTATCT
CTATCTACCT TCTTTATCTT TTCACATAA CTTGAATTC CTTACCTTT TGAACCTGA ACGAGTTTTA ACATGATAGA

490 CTATTCAGGA TTATGAAGTT TTTTCGATCG AAGATTCACT GGACGAAAGA AAGATAAAA GGTTCATTGA GCTCAGGAAG
GATAAGTCCT AATACTTCAA AAAGCTACGC TTCTAAGTGA CCTGCTTTCT TTCTATTTTC CCCAGTAACT CGAGTCCCTC

```

Fig. 4b

570	580	590	600	610	620	630	640
AGCTTACTGT	CTGCCCTTGAG	AACTTATGAA	CCATATGGAT	CCCTGGTTCA	ACAAATACGA	ATTCTGCTGC	TGGGTCCAAT
TCGAATGACA	GACGGAACTC	TTGAATACTT	GGTATACCTA	GGGACCAAGT	TGTTTATGCT	TAAGACGACG	ACCCAGGTTA
650	660	670	680	690	700	710	720
TGGAGCTGGG	AAGTCTAGCT	TTTTCAACTC	AGTGAGGTCT	GTTTTCCTCAAG	GGCATGTAAC	GCATCAGGCT	TTGGTGGGCA
ACCTCGACCC	TTCAGATCGA	AAAAGTTGAG	TCACTCCAGA	CAAAAAGGTC	CCGTACATTG	CGTAGTCCGA	AACCAACCCGT
730	740	750	760	770	780	790	800
CTAATACAA	TGGGATATCT	GAGAAGTATA	GGACATACTC	TATTAGAGAC	GGGAAAGATG	GCAAATACCT	GCCATTATT
GATTATGTTG	ACCTATAGA	CTCTTTCATAT	CCTGTATGAG	ATAATCTCTG	CCCTTCTTAC	CGTTTATGGA	CGGTAAATAA
810	820	830	840	850	860	870	880
CTGTGTGACT	CACCTGGGGCT	GAGTGAGAAA	GAAGGCGGCC	TGTGCATGGA	TGACATATCC	TACATCTTGA	ACGGTAACAT
GACACACTGA	GTGACCCCGA	CTCACTCTTTT	CTTCCGCGCG	ACACGTACCT	ACTGTATAGG	ATGTAGAACT	TGCCATTGTA
890	900	910	920	930	940	950	960
TCGTGATAGA	TACCAGTTTA	ATCCCATGGA	ATCAATCAAA	TTAAATCATC	ATGACTACAT	TGATTCCTCCA	TCGCTGAAGG
AGCACTATCT	ATGGTCAAA	TAGGGTACCT	TAGTTAGTTT	AATTTAGTAG	TACTGATGTA	ACTAAGGGGT	AGCGACTTCC
970	980	990	1000	1010	1020	1030	1040
ACAGAATTCA	TTGTGTGGCA	TTTGTAATTG	ATGCCAGCTC	TATTGAATAC	TTCTCCTCTC	AGATGATAGT	AAAGATCAAA
TGCTCTTAAGT	AACACACCCG	AAACATAAAC	TACGGTCCGAG	ATAACTTATG	AAGAGGAGAG	TCTACTATCA	TTTCTAGTTT
1050	1060	1070	1080	1090	1100	1110	1120
AGAATTCGAA	GGGAGTTGGT	AAACGCTGGT	GTGGTACATG	TGGCTTTGCT	CACATCATGTG	GATAGCATGG	ATCTGATTAC
TCTTAAGCTT	CCCTCAACCA	TTTGCGACCA	CACCATGTAC	ACCGAAACGA	GTGAGTACAC	CTATCGTACC	TAGACTAATG

0293274

Fig. 4c

1130	1140	1150	1160	1170	1180	1190	1200
AAAAGGTGAC	CTTATAGAAA	TAGAGAGATG	TGTGCCTGTG	AGGTCCAAGC	TAGAGGAAGT	CCAAAGAAAA	CTTGGATTGG
TTTTCCACTG	GAATATCTTT	ATCTCTCTAC	ACACGGACAC	TCCAGGTTCCG	ATCTCCTTCA	GGTTTCTTTT	GAACCTAAAC
1210	1220	1230	1240	1250	1260	1270	1280
CTCTTTCTGA	CATCTCGGTG	GTTAGCAATT	ATTCTCTCTGA	GTGGGAGCTG	GACCCCTGTAA	AGGATGTTCT	AATTCTTTCT
GAGAAAGACT	GTAGAGCCAC	CAATCGTTAA	TAAAGGAGACT	CACCCCTCGAC	CTGGGACATT	TCCTACAAGA	TTAAGAAAGA
1290	1300	1310	1320	1330	1340	1350	1360
GCTCTGAGAC	GAATGCTATG	GGCTGCAGAT	GACTTCTTAG	AGGATTGCCC	TTTTTGAGCAA	ATAGGGAATC	TAAGGGGAGGA
CGAGACTCTG	CTTACGATAC	CCGACGTCTA	CTGAAGAAATC	TCCTAAACGG	AAAACTCGTT	TATCCCTTAG	ATTCCCTCCT
1370	1380	1390	1400	1410	1420	1430	1440
AATTATCAAC	TGTGCACAAG	GAATAAATA	GATATGTGAA	AGGTTACAGT	AAATTTCTC	ACATCACAGA	AGATTAAAAAT
TTAATAGTTG	ACACGTGTTT	CTTTTTTTAT	CTATACACTT	TCCAAGTGCA	TTTAAAGGAG	TGTAGTGTCT	TCTAATTTTA
1450	1460	1470	1480	1490	1500	1510	1520
TCAGAAAGGA	GAACACACAG	ACCAAAGAGA	AGTAACTAAG	ACCAAAGGGA	TGTGTTTTAT	TAATGTCTAG	GATGAAGAAA
AGTCTTTTCT	CTTTTGTGTC	TGGTTTCTCT	TCATTGATTC	TGGTTTCCCT	ACACAAAATA	ATTACAGATC	CTACTTCTTT
1530	1540	1550	1560	1570	1580	1590	1600
TGCATAGAAC	ATTGTAGTAC	TTGTAAATAA	CTAGAAATAA	CATGATTTAG	TCATAATTGT	GAATAATAAT	AATAATTTT
ACGTATCTTG	TAAACATCATG	AACATTATTT	GATCTTTATT	GTAATAAATC	AGTATTAAAC	CTTTTATTAT	TTATTAAAAA
1610	1620	1630	1640	1650	1660		
CTTGGATTTA	TGTTCTGTAT	CTGTGAAAAA	ATAAATTCTT	TATAAAAAAA	AAAAAAAAAA	AAAAAAA 3'	
GAACCTAAAT	ACAAGACATA	GACACTTTTT	TATTTAAAGA	ATATTTTTTT	TTTTTTTTTT	TTTTTTTTT 5'	

0293274

Fig. 5

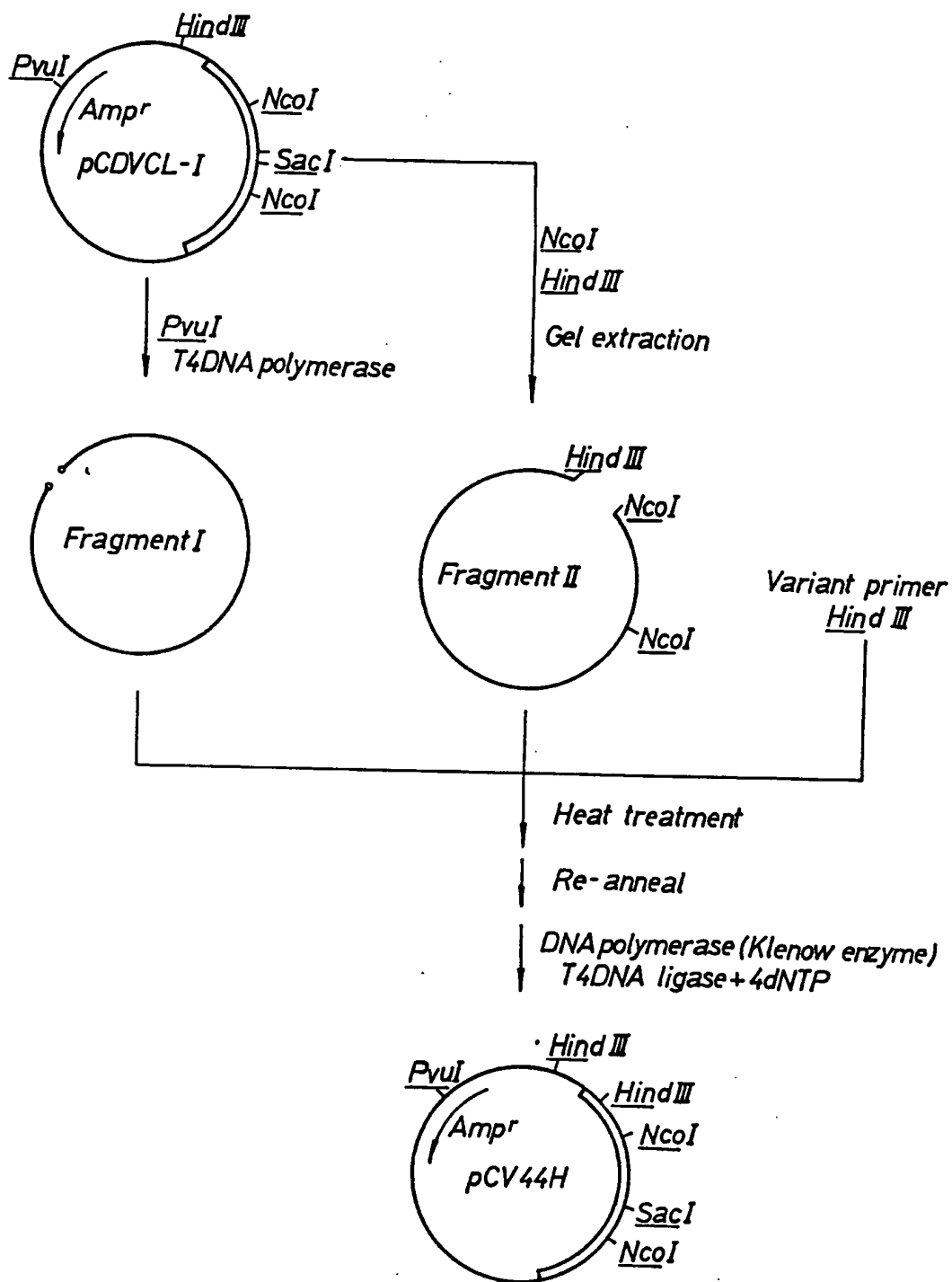


Fig. 6

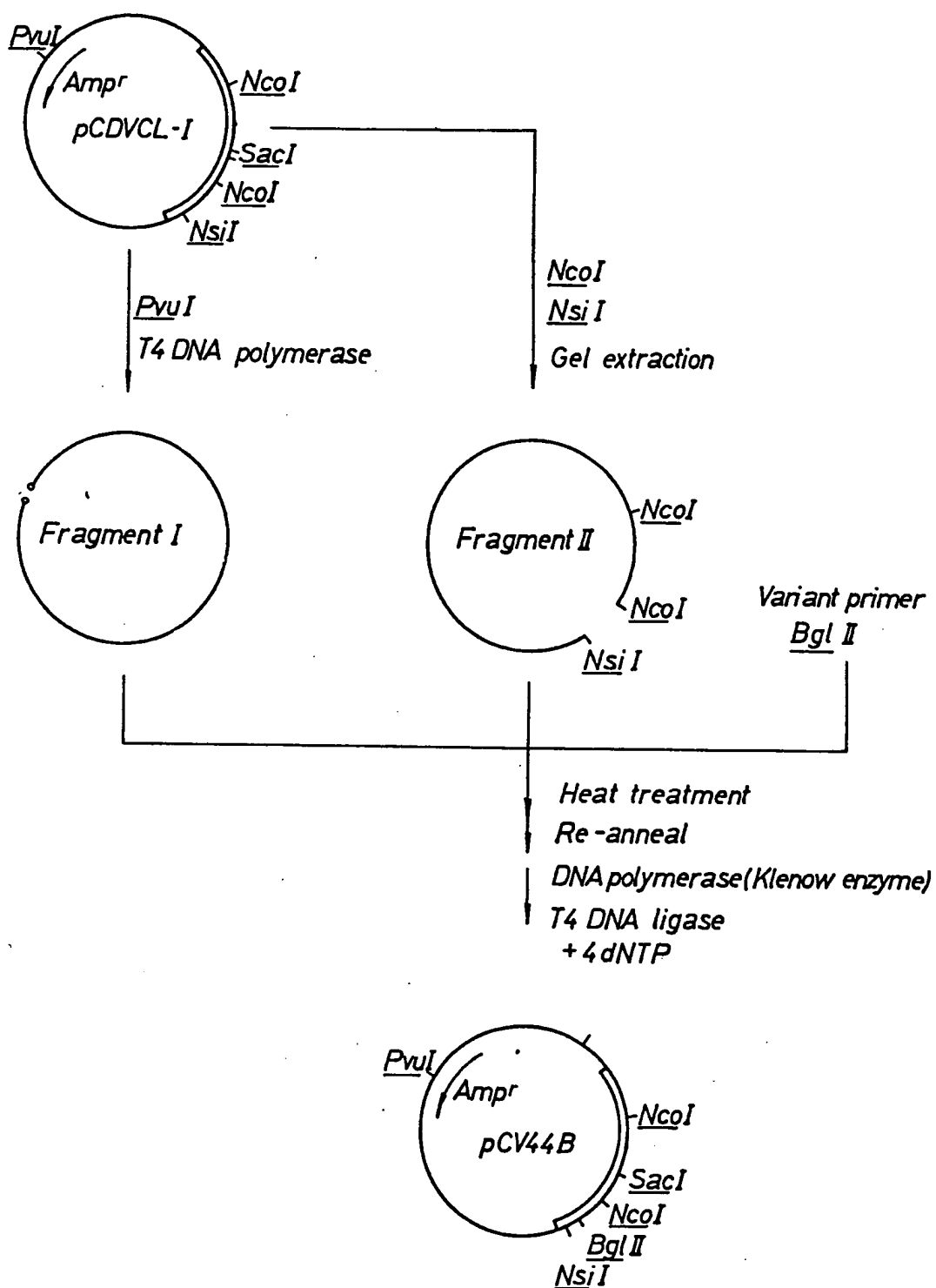
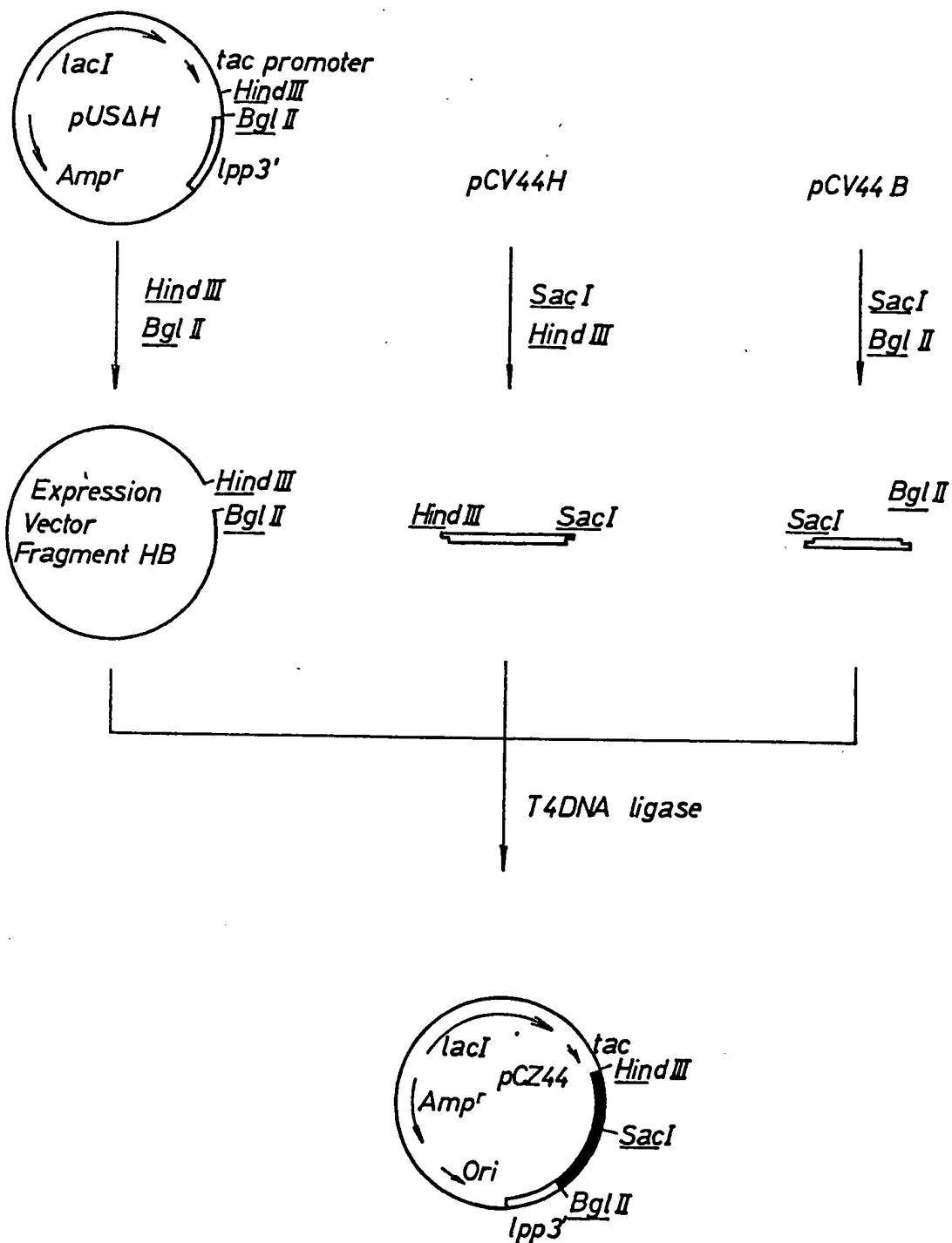


Fig.7







European Patent  
Office

## EUROPEAN SEARCH REPORT

Application Number

EP 88 40 0790

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)
D, Y	EP-A-0 190 972 (MITSUBISHI) * Whole document *	1-11	C 12 N 15/00 A 61 K 39/29
D, Y	PROC. NATL. ACAD. SCI. USA, vol. 80, March 1983, pages 1194-1198; R.A. YOUNG et al.: "Efficient isolation of genes by using antibody probes" * Whole article *	1-11	
A	CHEMICAL ABSTRACTS, vol. 105, no. 17, October 1986, page 527, no. 151110y, Columbus, Ohio, US; T. AKATSUKA et al.: "Non-A, non-B hepatitis related AN6520 Ag is a normal cellular protein mainly expressed in liver II" & J. MED. VIROL. 1986, 20(1), 43-56 * Whole abstract *	1	
A	BIOLOGICAL ABSTRACTS, vol. 80, 1985, no. 4871; J.-I. TOHMATSU et al.: "AN6520 antigen: an antigen purified from liver with non-A, non-B hepatitis" & J. MED. VIROL 15(4): 357-372. 1985. * Whole abstract *	1	
A	EP-A-0 066 296 (EISAI CO., LTD)		
A	EP-A-0 092 249 (EISAI CO., LTD)		
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 21-06-1988	Examiner SKELLY J.M.
<b>CATEGORY OF CITED DOCUMENTS</b> X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) Publication number: **0 293 274 B1**

(12)

## EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification :  
04.09.91 Bulletin 91/36

(51) Int. Cl.<sup>5</sup>: C12N 15/12, C12N 15/33

(21) Application number: 88400790.7

(22) Date of filing: 31.03.88

(54) DNA molecules encoding non-A, non-B hepatitis antigens, and their use in producing said antigens.

(30) Priority: 31.03.87 JP 78313/87  
04.06.87 JP 140588/87  
10.11.87 JP 283990/87

(43) Date of publication of application :  
30.11.88 Bulletin 88/48

(45) Publication of the grant of the patent :  
04.09.91 Bulletin 91/36

(84) Designated Contracting States :  
CH DE FR GB IT LI NL SE

(56) References cited :  
EP-A- 0 066 296  
EP-A- 0 092 249  
EP-A- 0 190 972  
PROC. NATL. ACAD. SCI. USA, vol. 80, March  
1983, pages 1194-1198; R.A. YOUNG et al.:  
"Efficient isolation of genes by using antibody  
probes"  
CHEMICAL ABSTRACTS, vol. 105, no. 17,  
October 1986, page 527, no. 151110y, Colum-  
bus, Ohio, US; T. AKATSUKA et al.: "Non-A,  
non-B hepatitis related AN6520 Ag is a normal  
cellular protein mainly expressed in liver II" &  
J. MED. VIROL 1986, 20(1), 43-56  
BIOLOGICAL ABSTRACTS, vol. 80, 1985, no.  
4871; J.-I. TOHMATSU et al.: "AN6520 antigen:  
an antigen purified from liver with non-A,  
non-B hepatitis" & J. MED. VIROL 15(4):  
357-372. 1985.

(73) Proprietor: MITSUBISHI KASEI  
CORPORATION  
5-2, Marunouchi 2-chome Chiyoda-ku  
Tokyo 100 (JP)

(72) Inventor: Takahashi, Kazuhiro 303 Corpo  
Omodaka  
29 Ichijo-Somedono-cho Sakyo-ku  
Kyoto City Kyoto Prefecture (JP)  
Inventor: Shibui, Tatsuro  
M606-11 San White 6-15-4 Minami-Naruse  
Machida City Tokyo (JP)  
Inventor: Kamizono, Michiru  
F202-24 San White 2-8-11 Tsuruma  
Machida City Tokyo (JP)  
Inventor: Matsui, Rie  
A211-23 San White 1-26-31 Ogawa  
Machida City Tokyo (JP)  
Inventor: Teranishi, Yutaka  
1-14-2 Takane  
Sagamihara City Kanagawa Prefecture (JP)  
Inventor: Nakanishi, Shigetada  
517-116 Iwakura-Nagatani-cho Sakyo-ku  
Kyoto City Kyoto Prefecture (JP)  
Inventor: Kitamura, Naomi  
2-272 Yakumo-Higashi-machi  
Moriguchi City Osaka Prefecture (JP)

(74) Representative: Gutmann, Ernest et al  
S.C. Ernest Gutmann - Yves Plasseraud 67,  
boulevard Haussmann  
F-75008 Paris (FR)

EP 0 293 274 B1

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

## Description

The present invention generally relates to the production of an antigen specific to non-A non-B hepatitis by recombinant DNA technology. More particularly, it relates to a DNA fragment coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis, an expression vector containing such a DNA fragment, a host transformed with such an expression vector, as well as a process for producing said antigen specific to non-A non-B hepatitis by culturing such a transformant.

Among viral hepatitises, the viral entities of hepatitis type A and type B have been found and, accordingly, it has now become possible to diagnose such diseases by immunological methods.

Still another type of hepatitis different from the types A and B, which is called non-A non-B type hepatitis, is said to be over 90% of post-transfusion hepatitis : refer to NIPPON RINSHO (Japan Clinic), 35, 2724 (1977); J. Biol. Med., 49, 243 (1976). The pathogenic virus of the non-A non-B type hepatitis, however, has not yet been identified. Only one fact which has already been established is potential infection of human hepatitis type non-A non-B virus to chimpanzee : refer to Lancet I, 459 (1978) ; *ibid.*, 463 (1978).

Many workers have done various investigations for searching an antigen-antibody system related to the non-A non-B hepatitis by using mainly sera from patients affected with the disease ; nevertheless, no definite system has been found. Under these circumstances, the diagnosis of non-A non-B hepatitis should inevitably be effected by so-called exclusion diagnosis : that is, whether or not the hepatitis of a patient is type A or type B or other hepatitis due to a virus known to cause hepatopathy, for example, CMV, HSV, EBV, etc., is first determined ; and if not, the patient's hepatitis is diagnosed as non-A non-B type. Thus, such a diagnosis of non-A non-B hepatitis will require much time and labor.

An antigenic protein specific to non-A non-B hepatitis and useful for the direct diagnoses of the hepatitis has been purified from human and chimpanzee hepatocytes affected with non-A non-B hepatitis, and a monoclonal antibody specific to the antigen and useful for the treatment of the non-A non-B hepatitis has also been proposed : refer to Japanese Patent Application Laying-open (KOKAI) Nos. 176856/86 and 56196/86.

A large amount of such an antigenic protein specific to non-A non-B hepatitis should be required when such a protein is to be employed, for example, as a diagnostic agent. However, it is not always appropriate to purify such a large amount of the antigenic protein from chimpanzee hepatocytes affected with non-A non-B hepatitis.

On the other hand, in order to detect a gene coding for a specific antigen of non-A non-B hepatitis by nucleic acid hybridization and, further, to produce such an antigen specific to non-A non-B hepatitis by the recombinant DNA technology, it is essential to obtain a gene fragment coding for the antigenic protein specific to the non-A non-B hepatitis.

The present inventors have made great efforts to produce such a specific antigenic protein in a large amount by genetic engineering techniques, and finally isolated a gene fragment coding for the antigenic protein specific to non-A non-B hepatitis, said gene fragment being useful for the production of such antigens. Further, the inventors have successfully constructed an expression vector containing said gene fragment. Thus, the present invention has now been attained.

It is an object of the invention to provide a DNA fragment which contains a base sequence coding for an antigen specifically occurring in a host cell affected with non-A non-B hepatitis or an antigenic protein specific to non-A non-B hepatitis having physiological activities equivalent to those of said specifically occurring antigen.

Another object of the invention is to provide an expression vector having said DNA fragment introduced thereinto at a cloning site downstream from a promoter of the vector.

A still another object of the invention is to provide a transformant obtained by transforming a host cell with said expression vector.

A further object of the invention is to provide a process for producing such an antigen specific to non-A non-B hepatitis by culturing said transformant.

Other objects and advantages of the present invention will be apparent from the following detailed description with reference to the attached drawings, in which :

- Figs. 1a-1e show the base sequence coding for an antigenic protein specific to non-A non-B hepatitis ;
- Fig. 2 shows the base sequence of a hybrid promoter Pac ;
- Figs. 3a-3c show the base sequence of a cDNA fragment obtained in Example 1 described hereinbelow, together with deduced amino acid sequence ;
- Figs. 4a-4c show the base sequence of cDNA containing the full length gene sequence of an antigenic protein specific to non-A non-B hepatitis, which cDNA was obtained in Example 2 described hereinbelow, the base sequence 57-1388 thereof coding for the antigenic protein specific to non-A non-B hepatitis ;
- Fig. 5 schematically illustrates the construction of a plasmid pCV44H ;
- Fig. 6 schematically illustrates the construction of a plasmid pCV44B ; and
- Fig. 7 schematically illustrates the construction of a plasmid pCZ44.

The present invention will be described in detail hereinbelow.

According to one aspect of the invention, a DNA fragment is provided which contains a base sequence coding for an antigenic protein occurring specifically in hepatocytes affected with non-A non-B hepatitis.

Such a DNA fragment of the invention may be prepared in the following manner.

First, a liver tissue specimen derived from a human or chimpanzee individual affected with non-A non-B hepatitis is homogenized in an aqueous solution of guanidinium thiocyanate and then subjected to cesium chloride equilibrium density gradient centrifugation according to Chirgwin *et al.* method (Biochemistry, 18, 5294-5299 (1979)) to separate total RNA as a precipitate. After separation, the total RNA is purified by phenol extraction and ethanol precipitation.

"Individuals affected with non-A non-B hepatitis" used as sources of liver tissue specimens in the invention may include those affected with so-called type D hepatitis, which has recently been named.

It is known that mRNA of an antigen gene generally has a poly-A chain. Thus, the total RNA is subjected to oligo(dT) cellulose column chromatography in a conventional manner and poly(A)-containing RNA (poly A<sup>+</sup> RNA) is isolated as mRNA material.

A cDNA library corresponding to the poly A<sup>+</sup> RNA is then obtained from the mRNA material according to the random primer method (Y. Ebina *et al.*, Cell, 40, 747-758 (1980)) : Thus, a number of DNAs complementary to the mRNA material are randomly synthesized using any primer of e.g. about 6 bases and a reverse transcriptase.

The cDNA is methylated with a DNA methylase, e.g. EcoRI methylase, to protect cleavage sites present in the cDNA capable of being cleaved by a corresponding restriction enzyme, e.g. EcoRI. A DNA linker containing the corresponding restriction enzyme cleavage sites at both ends, e.g. EcoRI linker (CGAATTCG), is added to the methylated cDNA and, then, this cDNA is digested with the restriction enzyme, e.g. EcoRI.

The digested cDNA is then cloned into a cloning vector such as a plasmid or a  $\lambda$  phage. For example, the cDNA may be introduced into EcoRI site of  $\lambda$ gt 11 DNA, which is an expression cloning vector : refer to R.A. Young *et al.*, Proc. Natl. Acad. Sci. U.S.A., 80, 1194-1198 (1983). The cDNA will be inserted into the  $\beta$ -gal gene on the  $\lambda$ gt 11 phage. Thus, expression of the cDNA can be easily verified by the production of a fused protein with  $\beta$ -galactosidase due to induction of the expression by the lactose operon promoter of said phage when E. coli transfected with said phage is cultured in a medium containing IPTG (isopropylthio- $\beta$ -D-galactopyranoside).

The  $\lambda$ gt 11 phage incorporating the cDNA is then introduced into E. coli by Tomizawa *et al.* method in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). The thus transfected microorganism is cultured in an IPTG-containing medium.

The thus formed plaques can be easily selected by an immunological screening method using a monoclonal antibody specifically directed to non-A non-B hepatitis to obtain a desired cDNA. Such a monoclonal antibody which can be used in the immunological screening method may be prepared according to the methods described in Japanese Patent Application Laying-open Nos. 56196/86 and 91328/88. The screening methods used may include the western blotting technique described in these applications.

The plaques positive in the immunological screening test are selected to proliferate the phage by Tomizawa *et al.* method. DNA is purified from the grown phage by T. Maniatis *et al.* method in "Molecular Cloning", Cold Spring Harbor Laboratory, pp. 85 *et seq.* (1982), and cleaved with a suitable restriction enzyme such as EcoRI. The thus purified and digested DNA fragments can be used to determine the base sequence of a desired cDNA segment according to Maxam and Gilbert method in Methods in Enzymology, 65, 499-560 (1980) ; or alternatively, after further cloning the DNA fragments into M13 phage, the base sequence of such a desired cDNA segment can be determined according to the dideoxy method : Sanger *et al.*, Proc. Natl. Acad. Sci. U.S.A., 74, 5463 (1977).

Thus, a cDNA fragment coding for an antigen specific to non-A non-B hepatitis can be obtained. However, such a DNA fragment may usually be only a portion of the gene coding for the non-A non-B hepatitis-specific antigen.

A full length cDNA coding for such a non-A non-B hepatitis-specific antigen may be obtained in the following manner.

Poly A<sup>+</sup>-mRNA is isolated and purified in a manner similar to that described above. From the poly A<sup>+</sup>-mRNA a cDNA library is obtained according to Okayama-Berg vector-primer method : Molecular and Cellular Biology, 2, 161-170 (1982).

A plasmid containing such a cDNA thus prepared is used to transform E. coli by any conventional method, for instance, the method D. Hanahan : J. Mol. Biol., 166, 557 (1983). The transformant ampicillin-resistant strains are collected and screened by the colony hybridization method using the aforementioned DNA fragment as a probe. Such a probe may preferably be prepared by either the strepto-avidin method, or the nick translation method using photobiotinnucleic acids and <sup>32</sup>P-nucleic acids.

The thus selected colonies containing a cDNA clone are cultured. Plasmid DNA is obtained from the cul-

5 tured colony according to Birnboim et al. method (Nucleic Acid Res., 7, 1513 (1979)) and digested with a suitable restriction enzyme. The base sequence of a desired full-length cDNA segment is then determined according to the aforementioned Maxam and Gilbert method or, alternatively, after further cloning the digested DNA into M13 phage or pVC12 plasmid, such a base sequence is determined according to the above described Sanger et al. dideoxy method.

The base sequence of the full length DNA coding for an antigen specific to non-A non-B hepatitis is shown in Fig. 1, in which the symbol "-" just under the base sequence represents a corresponding base complementary to the respective base described just above each of the symbols.

10 Of course, DNA fragments which can be employed in the invention do not necessarily contain the same base sequence as shown in Fig. 1, but those DNA fragments in which a part of said base sequence shown in Fig. 1 has been substituted by at least one different base or deleted therefrom and those DNA fragments in which one or more additional bases have been added to the base sequence of Fig. 1 may also be included herein provided that such different DNA fragments may code for substances having physiological activities equivalent to those of the non-A non-B hepatitis-specific antigens encoded by the base sequence of Fig. 1.

15 According to another aspect of the invention, an expression vector is provided in which the aforementioned DNA fragment of the invention is inserted into a cloning site downstream from a promoter of this vector.

The expression vector of the invention contains a promoter in a position capable of controlling the transcription of a DNA fragment coding for a non-A non-B hepatitis-specific antigen obtained by the aforementioned method. The promoters used in the invention may be any promoter capable of expressing the DNA fragment in a host, and preferably of controlling the transcription of the fragment.

20 When a host used is a microorganism such as Escherichia coli, Bacillus subtilis, etc., the expression vector of the invention may preferably comprise a promoter, a ribosome binding sequence, a gene for a non-A non-B hepatitis-specific antigen, a transcription termination factor, and a gene controlling the promoter.

25 The promoter used may include those derived from E. coli, phage, etc., for example, tryptophan synthase operon (trp), lactose operon (lac), lipoprotein (lpp), recA, lambda phage P<sub>L</sub>, P<sub>R</sub>, T5 early gene P<sub>25</sub>, P<sub>28</sub> promoter, which may also be prepared by chemical synthesis. Also included herein are hybrid promoters such as tac (trp : lac), trc (trp : lac) and Pac (phage : E. coli) shown in Fig. 2.

The ribosome binding sequence may be derived from E. coli, phage, etc., but preferably may be those synthetically prepared, for example, those containing a consensus sequence such as

AGGAGGTTTAA.  
SD sequence

35 The gene for a non-A non-B hepatitis-specific antigen may be directly employed without any modification. Preferably, an unnecessary base sequence (non-coding region) may be deleted by site-directed mutagenesis: BIO TECHNOLOGY, July, 636-639 (1984).

A transcription termination factor may not always be required in the expression vector of the invention. Preferably, the instant vector may contain a p-independent terminator, for example, lpp terminator, trp operon terminator, ribosomal RNA gene terminator, etc.

40 The expression vector may be derived from any conventional plasmid. Preferably, it may be derived from such a plasmid as replicating itself in E. coli or Bacillus subtilis, for example, pBR322- or pUB110-derived plasmid.

45 Desirably, these factors required for expression are arranged in the expression plasmid in the order of the promoter, the SD sequence, the structural gene of a non-A non-B hepatitis-specific antigen, and the transcription termination factor from 5' to 3'. A repressor gene required to control the transcription, a marker gene such as drug-resistant gene, and a plasmid replication origin may be arranged in any order in the expression vector.

The expression vector of the invention may be introduced into a host by any conventional method for transformation of E. coli, e.g., one described in Molecular Cloning, 250-253 (1982), or of Bacillus subtilis, e.g., one described in Molec. Gen. Genet., 168, 111-115 (1979) or Proc. Nat. Acad. Sci. U.S.A., 44, 1072-1078 (1958).

50 The resulting transformant may be cultured in any conventional medium, e.g. one described in Molecular Cloning, 68-73, (1972), at a temperature in the range of 28 to 42°C in both cases of E. coli and Bacillus subtilis. Preferably, it may be cultured at a temperature in the range of 28 to 30°C where no expression of heat shock proteins may be induced.

55 The desired protein thus produced may be easily purified from the host in conventional procedures. For example, the host cell may be crushed by lysozyme-surfactant or ultra-sonication, and the insoluble fractions which contain the desired non-A non-B hepatitis-specific antigen may be then collected by centrifugation, solubilized in a surfactant such as 0.01% SDS, and subjected to column chromatography using a monoclonal anti-

body (Japanese Patent Application Laying-open (KOKAI) Nos. 56196/86 and 176856/86.

When an eukaryotic cell such as an animal cell is employed as a host, the expression vector of the invention is preferably as follows :

5 The promoters used in the vector of the invention for the expression in eukaryotic cells may herein include SV40 early and late promoters ; promoters of apolipoprotein E and A-I genes ; promoter of heat shock protein gene (Proc. Natl. Acad. Sci. U.S.A., 78, 7038-7042 (1981)) ; promoter of metallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6511-6515 (1980)) ; HSV TK promoter ; adenovirus promoter, such as Ad2 major late promoter (Ad2 MLP) ; LTR (long terminal repeat) of retrovirus ; etc. SV40 promoter and promoter of metallothionein gene are preferred.

10 The expression vector of the invention may contain a splice sequence comprising 5' splice junction donor site, an Intron and 3' splice junction acceptor site. A common base sequence is found at all the splice junction sites (exonintron junction sites) ; so-called GT/AG rule that any intron region always starts from two bases GT at the donor site and terminates at two bases AG of the acceptor site has been established.

The expression vector of the invention may contain one or more splice sequences as mentioned just above. 15 Such splice sequences may be positioned upstream or downstream of the structural gene for a non-A non-B hepatitis-specific antigen.

Illustrative examples of such splice sequences may include those DNA sequences found in exons 2 and 3 of rabbit  $\beta$ -globin gene (Science, 26, 339 (1979)) and mouse methallothionein-I gene containing the promoter, exons 1, 2 and 3 and Introns A and B of methallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 77, 6513 (1980)). 20 The 5' and 3' splice sites may be derived from the same or different gene ; for example, a sequence in which 5' splice site contained in adenovirus DNA is linked to 3' splice site derived from the gene of Ig variable region can be employed.

The expression vector of the invention also contains a polyadenylation site downstream from the structural gene of a non-A non-B hepatitis-specific antigen. Illustrative examples of the polyadenylation sites may include 25 those derived from SV40 DNA,  $\beta$ -globin gene or methallothionein gene. A combined site of the polyadenylation sites of  $\beta$ -globin gene and SV40 DNA may be employed in the invention.

The expression vector of the invention may also contain a dominant selective marker permitting the selection of transformants. Selective markers which can be used herein may include DHFR gene imparting MTX (methotrexate) resistance to a host ; tk gene of herpes simplex virus (HSV) which permits selection of tk<sup>-</sup> strains 30 transformed therewith in HAT medium ; the gene for aminoglycoside 3'-phosphotransferase from *E. coli* transposon Tn5, which imparts to a host the resistance against 3'-deoxystreptamine antibiotic G418 ; bovine papilloma virus gene permitting morphological discrimination by piled up growth ; and aprt gene.

Alternatively, animal cells transformed with the expression vector of the invention may be selected by the cotransformation even though no selective marker is present in the vector. For this purpose, an animal cell is 35 cotransformed with both the expression vector and a plasmid or other DNA containing a gene for such a selective marker and selected by a phenotypic trait of the gene.

Advantageously, the expression vectors may also contain a plasmid fragment having an origin of replication derived from a bacterium such as *E. coli*, since such vectors can be cloned in bacteria. Such plasmids may include pBR322, pBR327, pML, etc.

40 Illustrative examples of plasmid vectors used as sources of the expression vectors according to the invention may include pKCR (Proc. Natl. Acad. Sci. U.S.A., 78, 1528 (1981)), which contains SV40 early promoter, the splice sequence and polyadenylation site derived from rabbit  $\beta$ -globin gene, the polyadenylation site from SV40 early region, and the origin of replication and ampicillin resistant gene from pBR322 ; pKCR H2 (Nature, 307, 605 (1984)), in which the pBR322 portion of pKCR has been substituted by pBR327 fragment and the 45 *EcoRI* site present in the exon 3 of rabbit  $\beta$ -globin gene has been converted into *HindIII* site ; and pBPV MT1 containing BPV gene and methallothionein gene (Proc. Natl. Acad. Sci. U.S.A., 80, 398 (1983)).

Animal cells transformed with the expression vector of the invention may include CHO cells, COS cells, and mouse L cells, C127 cells and FM3A cells.

The introduction of the expression vector of the invention into an animal cell may be carried out by transfection, microinjection, etc. Most often, the transfection may employ  $\text{CaPO}_4$  ; Virology, 52, 458-467 (1973). 50

Animal cells transformed by introducing the expression vector of the invention may be cultured in a suspension or solid medium by conventional methods. The culture medium used is most often MEM, RPMI1640, etc.

55 Proteins produced in the transformed animal cells can be separated and purified in the almost same manner as in the case of microorganisms aforementioned.

As stated, the invention provides a transformant cell obtained by introducing the expression vector of the invention into a host cell.

Also provided according to the invention is a process for producing a non-A non-B hepatitis-specific antigen

comprising culturing said transformant and collecting the produced and accumulated antigen.

As stated previously, a large amount of an antigenic protein specific to non-A non-B hepatitis is required when such a protein is to be utilized as a direct diagnostic agent. According to the present invention, such an antigenic protein can be produced with a low cost and a large scale without use of infected chimpanzee hepatocytes. Prior to the present invention, it has been difficult to obtain such a large amount of a non-A non-B hepatitis-specific antigenic protein from hepatocytes of chimpanzees affected with non-A non-B hepatitis.

Further, the DNA fragment coding for an antigenic protein of non-A non-B hepatitis virus according to the present invention will be useful as a probe for detecting the gene of said antigenic protein by nucleic acid hybridization.

## EXAMPLES

The following examples will be given by way of illustration but these examples in no way limit the scope of the invention without departing the concept thereof.

### EXAMPLE 1 : Preparation of cDNA Fragment Coding for Antigenic Protein Specific to Non-A Non-B Hepatitis

Poly(A)-containing RNA was prepared from chimpanzee liver according to the guanidine thiocyanate-lithium chloride method : Cathala et al., DNA, 2, 329 (1983).

The infected liver (5 g) was taken out from a chimpanzee affected with non-A non-B hepatitis and immediately frozen by liquid nitrogen. The frozen liver was added into a Waring blender together with liquid nitrogen and ground at 3,000 rpm for 2 minutes. The ground liver specimen was further ground by a Teflon homogenizer at 5 rpm in 100 ml of a solution : 5 M guanidine thiocyanate, 10 mM EDTA, 50 mM Tris-HCl (pH 7), 8% (v/v)  $\beta$ -mercaptoethanol. The thus solubilized material (20 ml) was slowly placed on 5.7 M CsCl solution (10 ml) contained in a centrifuge tube and centrifuged at 27,000 rpm for 20 hours in Hitachi RPS 28-2 rotor. The thus precipitated RNA was collected and dissolved in 10 ml of a solution : 0.1% sodium laurylsulfate, 1 mM EDTA, 10 mM Tris-HCl (pH 7.5). The RNA was extracted with phenol-chloroform and recovered by ethanol precipitation.

The thus obtained RNA (about 3.95 mg) was dissolved in 1 ml of a solution : 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. The solution was incubated at 65°C for 5 minutes, and 5 M NaCl (0.1 ml) was added. The resulting mixture was subjected to chromatography on an oligo(dT) cellulose column (column volume of 0.5 ml, P-L Biochemical). The thus adsorbed poly(A)-containing mRNA was eluted with a solution : 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. There was obtained about 100  $\mu$ g of poly(A)-containing mRNA.

The thus obtained poly(A)<sup>+</sup> mRNA (10  $\mu$ g) was dissolved in 50  $\mu$ l of RT buffer : 20 mM Tris-HCl (pH 8.8), 0.1 M KCl, 12 mM MgCl<sub>2</sub>, 2 mM MnCl<sub>2</sub>. To this solution, there was added 8  $\mu$ g of random primer d(N)<sub>6</sub> (P-L Biochemical). The resulting mixture was heated at 95°C for 3 minutes to denature the materials, which was then cooled gradually to room temperature to anneal the random primer with the mRNA. To the annealed mixture, there were added 10 mM 4NTP (10  $\mu$ l) and reverse transcriptase (225 units) from TAKARA SHUZO (Japan), and then water was added so as to make the total volume of the mixture to 100  $\mu$ l. Reaction was allowed to proceed at 42°C for one hour.

To the reaction mixture (50  $\mu$ l), there were added 10 mM NAD (2  $\mu$ l), 10 mM 4dNTP (10  $\mu$ l), RNase H (5 units), *E. coli* ligase (1 unit), *E. coli* DNA polymerase I (6.3 units), and 10x T4 DNA ligase buffer (10  $\mu$ l ; 0.1 M Tris-HCl, pH 7.5, 0.1 M DTT, 60 mM MgCl<sub>2</sub>) to make the total volume to 100  $\mu$ l. The mixture was allowed to react at 37°C for one hour to synthesize a double stranded DNA.

The thus obtained double stranded DNA was extracted with an equal volume of water-saturated phenol. Phenol in the aqueous layer was removed with the aid of ether followed by ethanol precipitation. The precipitate thus obtained was dissolved in 50  $\mu$ l of water, and 10x T4 DNA polymerase buffer (10  $\mu$ l ; 0.33 M Tris-acetic acid, pH 7.9, 0.66 M potassium acetate, 0.1 M magnesium acetate, 5 mM DTT), 10 mM 4dNTP (10  $\mu$ l), and T4 DNA polymerase (6 units) were added to make the total volume to 100  $\mu$ l. The mixture was reacted at 37°C for one hour. There was obtained a double stranded DNA having blunt ends, which was then extracted with phenol to remove proteins and purified by ethanol precipitation as described above. The thus purified DNA was then air dried.

To the purified DNA, there were added 50 mM Tris-HCl (pH 7.5), 1 mM Na<sub>2</sub>EDTA, 5 mM DTT (20  $\mu$ l), 100  $\mu$ M S-adenosyl-L-methionine (2  $\mu$ l), and 1.8 mg/ml *Eco*RI methylase (0.2  $\mu$ l). Reaction was effected at 37°C for 15 minutes, whereby methylating the *Eco*RI restriction enzyme cleavage site on the DNA fragment. The reaction mixture was then heated at 70°C for 15 minutes to deactivate the enzyme.

To the reaction mixture, there was added 3'-phosphorylated *Eco*RI linker (GGAATTC) in an amount of

100 molecules thereof per molecule of the synthetic DNA. There were further added 10x T4 DNA ligase buffer (5  $\mu$ l ; 0.5 M Tris-HCl, pH 7.5, 60 mM MgCl<sub>2</sub>, 10 mM DTT), 0.1 M ATP (5  $\mu$ l), and T4 DNA ligase (5 units) to make the total volume to 50  $\mu$ l. The resulting reaction mixture was reacted at 4°C for 16 hours followed by heating at 70°C for 10 minutes to deactivate the enzyme. Then, 10x EcoRI buffer (10  $\mu$ l ; 15 M Tris-HCl, pH 7.5, 0.5 M NaCl, 60 mM MgCl<sub>2</sub>), and EcoRI (100 units) were added to make the total volume to 100  $\mu$ l, and the reaction mixture was reacted at 37°C for 2 hours to cut the linker. The reaction mixture was passed through Bio Gel A-50 (0.2 cm x 32 cm, Bio Rad). Elution was effected by a buffer : 10 mM Tris-HCl (pH, 7.5), 6 mM MgCl<sub>2</sub>. Excess EcoRI linker was removed and, thus, a double stranded cDNA having EcoRI sites at both ends thereof was purified.

To the thus obtained double stranded cDNA fragment having EcoRI sites at both ends, there were added  $\lambda$ gt 11 DNA (10  $\mu$ g) cleaved with EcoRI, 10x T4 DNA ligase buffer (10  $\mu$ l) as described above, 0.1 M ATP (10  $\mu$ l), and T4 DNA ligase (10 units) to make the total volume to 100  $\mu$ l. The mixture was reacted at 4°C for 16 hours. Thus, said double stranded cDNA fragment was inserted into  $\lambda$ gt 11 DNA.

The  $\lambda$  phage packaging kit (PROMEGA, Biotech) was used to introduce said DNA into  $\lambda$  phage particle. The procedures for packaging were effected according to the instructions of the kit.

The  $\lambda$ gt 11 phage having said DNA packaged therein was used to transfect *E. coli* strain Y1090 to form plaques according to the conventional Tomizawa *et al.* methods described in "Experimental Procedures for Bacteriophages", pp. 99-174, published May 30, 1970 by Iwanami Shoten (Japan). Among about 200,000 plaques, one positive clone was selected by immunological screening as described hereinbelow. A monoclonal antibody used in the immunological screening was prepared by the method described in Japanese Patent Application Laying-open (KOKAI) No. 91328/88.

*E. coli* Y1090 (R.A. Young *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 80, 1194-1198 (1983), which had been transfected with  $\lambda$ gt 11, was inoculated in a petri dish together with soft agar held at 42°C. The transfected cell was allowed to stand at 42°C for 5 hours. A nitrocellulose filter (S & S, BA-83, pore size of 0.2  $\mu$ m) containing 10 mM IPTG was placed on the cell in the dish and incubation was effected at 37°C for 3-4 hours. This nitrocellulose filter was lightly rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed in the TBS buffer (400 ml) containing 3% gelatine and shaken at 40°C for one hour. Thus, the nitrocellulose filter was blocked. Then, a monoclonal antibody (OD<sub>280</sub> = 4.3) directed to a non-A non-B hepatitis-specific antigen was added to TBS buffer containing 1% gelatine with a dilution of 1/400. This mixture was put into a vinyl bag together with the filter in a proportion of 2 ml of the mixture per filter, and reaction was allowed to proceed at room temperature for 16 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes. A labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad) was added to TBS buffer containing 1% gelatine with a dilution of 1/1,000. This mixture and the filter were put into a vinyl bag with a proportion of 2 ml of the mixture per filter. Reaction was allowed to proceed at room temperature for 2 hours. The reaction mixture was three times washed with TBS buffer (400 ml) containing 0.05% Tween 20 over 10 minutes, in the same manner as described above. Color development was effected by dipping the filter and 4-chloro-1-naphthol (12 mg, Bio Rad) into 20 ml of TBS buffer containing hydrogen peroxide. After completion of the color development, the filter was thoroughly washed with water and put into a vinyl bag containing water. The bag was stored in a dark and cold place.

Thus, one positive plaque was obtained. The plaque was three times subjected to single plaque isolation. In each time, immunological screening was effected in the same manner as described above, verifying that the plaque was in fact positive.

The phage was then cultured in a large scale to purify the DNA in the following manner : First, *E. coli* Y1090 was cultured overnight in 10 ml of NZ medium prepared by adding NZ amine (10 g), NaCl (5 g) and 5 mM MgCl<sub>2</sub> to one liter of water followed by adjusting the pH to 7.2. The culture (1 ml) was transfected with the phage, with the m.o.i. (multiplicity of infection) being 0.1. The transfected culture was allowed to stand at 37°C for 10 minutes and then transferred to one liter of NZ medium. Shaking culture was effected at 37°C for 7-8 hours until the cells were lysed. Chloroform (5 ml) was added to the culture and shaking was continued for additional 30 minutes. The culture was subjected to centrifugation at 6,500 rpm for 10 minutes to remove cell debris.

NaCl (29 g) and polyethylene glycol (70 g) were added to and thoroughly dissolved in the obtained supernatant, and the solution was allowed to stand at 4°C overnight. The precipitate was collected by centrifugation at 6,500 rpm for 20 minutes, drained thoroughly, and dissolved in 20 ml of TM buffer : 10 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub>. DNase I and RNase A were added to the solution, both with a concentration of 10  $\mu$ g/ml, and the reaction was effected at 37°C for one hour. Chloroform (20 ml) was then added to the reaction mixture and stirred ; thus, polyethylene glycol was distributed in the chloroform layer which was then separated from the aqueous layer. This aqueous layer was ultra-centrifuged at 28,000 rpm for 60 minutes. Thus, a pellet of phage particles was obtained.

This pellet was dissolved in TM buffer (1 ml) and subjected to CsCl density gradient centrifugation at 33,000



rpm for 20 hours. The resultant fraction containing the phage particles ( $\rho = 1.45-1.50$ ) was dialyzed overnight against TM buffer. Proteinase K was added to the dialyzate in an amount of 100  $\mu\text{g/ml}$  and reaction was effected at 37°C for one hour. Thereafter, an equal volume of water-saturated phenol was added and phenol-extraction was gently effected. After centrifugation at 6,500 rpm for 10 minutes, the aqueous layer was removed, put into a dialysis tube, and dialyzed overnight against water at 4°C. Thus, about 5 mg of DNA was obtained.

Cleavage reaction of this DNA (100  $\mu\text{g}$ ) with EcoRI (100 units) in the aforementioned buffer (100  $\mu\text{l}$ ) at 37°C revealed that two cDNA segments of 390 bp and 345 bp were inserted into the phage DNA.

These two EcoRI fragments were re-cloned into EcoRI site of a cloning vector pUC 119. Base sequences of these DNA fragments were determined by the dideoxy method using commercially available primers CAG-GAAACAGCTATGAC and AGTCACGACGTTGTA, respectively. The base sequence of the linking portion between these two DNA fragments was similarly determined by cutting this cDNA fragment at BamHI and EcoRV sites present therein with corresponding specific restriction enzymes, inserting the resulting BamHI-EcoRV DNA fragment between BamHI and SmaI sites of the plasmid pUC 119, and sequencing the fragment by the dideoxy method.

The base sequence of said cDNA fragment is shown in Fig. 3. This was a partial cDNA fragment of a gene coding for an antigenic protein specific to non-A non-B hepatitis.

#### EXAMPLE 2 : Preparation of cDNA Containing the Full Length Gene Sequence

Messenger RNA was prepared as described in Example 1 and cDNA was synthesized using Okayama vector according to the conventional method described in Molecular Cloning, p. 211 *et seq.* The procedures used to synthesize cDNA were as follows :

To 300  $\mu\text{l}$  of a solution (10 mM Tris-HCl, pH 7.5, 6 mM  $\text{MgCl}_2$ , 10 mM NaCl), there were added 400  $\mu\text{g}$  of pCDV 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 500 units of KpnI (TAKARA SHUZO, Japan), all restriction enzymes used hereinafter having been manufactured by TAKARA SHUZO (Japan) unless otherwise noted. Reaction was effected at 37°C for 6 hours to cut the plasmid at KpnI site therein. After phenol-chloroform extraction, ethanol precipitation was effected to recover DNA.

The DNA (about 200  $\mu\text{g}$ ) cleaved with KpnI was added to 200  $\mu\text{l}$  of a solution which was obtained by adding dTTP in a concentration of 0.25 mM to a buffer (TdT buffer) : 40 mM sodium cacodylate, 30 mM Tris-HCl (pH 8.8), 1 mM  $\text{CaCl}_2$ , 0.1 mM dithiothreitol (DTT). Further, 81 units of terminal deoxynucleotidyl transferase (TdT, manufactured by P-L Biochemicals) was also added. Reaction was effected at 37°C for 11 minutes. Thus, a poly(dT) chain (about 67 deoxythymidylic acid residues) was added to the 3' end at the KpnI-cleaved site of pCDV 1. After phenol-chloroform extraction and ethanol-precipitation, about 100  $\mu\text{g}$  of pCDV 1 DNA to which poly(dT) chain had been added was recovered from the reaction mixture.

The thus obtained DNA was added to 150  $\mu\text{l}$  of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM  $\text{MgCl}_2$ , 100 mM NaCl), and HpaI (380 units) was also added, followed by reaction at 37°C for 2 hours. The reaction mixture was subjected to electrophoresis on agarose gel to separate and recover about 3.1 Kbp DNA fragment. Thus, there was obtained about 60  $\mu\text{g}$  of poly(dT)-containing pCDV 1.

The thus obtained DNA was dissolved in 500  $\mu\text{l}$  of a solution (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), incubated at 65°C for 5 minutes, and cooled on ice. After adding 5 M NaCl (50  $\mu\text{l}$ ), the mixture was subjected to chromatography on oligo(dA) cellulose column (Colaborative Research). DNA having a poly(dT) chain of sufficient length was adsorbed on the column and eluted with a solution : 10 mM Tris-HCl (pH 8.0), 1 mM EDTA. Thus, there was obtained 27  $\mu\text{g}$  of pCDV 1 to which poly(dT) chain had been added, abbreviated hereinafter as vector primer.

A linker DNA was prepared in the following manner : To 200  $\mu\text{l}$  of a solution (10 mM Tris-HCl, pH 7.5, 6 mM  $\text{MgCl}_2$ , 50 mM NaCl), there were added about 14  $\mu\text{g}$  of pL 1 (Okayama and Berg, Mol. Cell. Biol., 3, 280 (1983)) and 50 units of PstI. Reaction was effected at 37°C for 4 hours to cut the pL 1 DNA at PstI site. Phenol-chloroform extraction and ethanol precipitation of the reaction product gave about 13  $\mu\text{g}$  of pL 1 DNA cleaved at PstI site.

The thus obtained DNA (about 13  $\mu\text{g}$ ) was added to 50  $\mu\text{l}$  of the TdT buffer containing dGTP at a final concentration of 0.25 mM, and 54 units of TdT (P-L Biochemicals) was also added. The mixture was incubated at 37°C for 13 minutes to add a (dG) chain (about 14 deoxyguanylic acid residues) to the 3' end at the PstI-cleaved site of pL 1. After phenol-chloroform extraction, DNA was recovered by ethanol precipitation.

The thus obtained DNA was added to 100  $\mu\text{l}$  of a buffer (10 mM Tris-HCl, pH 7.5, 6 mM  $\text{MgCl}_2$ , 60 mM NaCl), and 80 units of HindIII was also added. The mixture was incubated at 37°C for 3 hours to cut the pL 1 DNA at HindIII site. The reaction product was fractionated by agarose gel electrophoresis. About 0.5 Kb DNA fragment was recovered by the DEAE paper method : Dretzen *et al.*, Anal. Biochem., 112, 295 (1981). Thus, there was obtained an oligo(dG) chain-containing linker DNA, hereinafter abbreviated simply as linker DNA.

The aforementioned poly(A)<sup>+</sup> RNA (about 2 µg) prepared in the same manner as in Example 1 and the vector primer (about 1.4 µg) were dissolved in 22.3 µl of a solution : 50 mM Tris-HCl (pH 8.3), 8 mM MgCl<sub>2</sub>, 30 mM KCl, 0.3 mM DTT, 2 mM dNTP (dATP, dTTP, dGTP and dCTP) and 10 units of ribonuclease inhibitor (P-L Biochemicals). To the solution, there was added 10 units of reverse transcriptase manufactured by SEIKAGAKU KOGYO (Japan). Incubation was effected at 37°C for 40 minutes to synthesize a DNA complementary to the mRNA. After phenol-chloroform extraction and ethanol precipitation, the vector primer DNA to which a double stranded RNA-DNA had been added was recovered.

The thus obtained vector primer DNA containing RNA-DNA double stranded chain was dissolved in 20 µl of TdT buffer containing 60 µM dCTP and 0.2 µg poly(A). After adding 14 units of TdT (P-L Biochemical), the mixture was incubated at 37°C for 8 hours to add a (dC) chain of 12 deoxycytidylic acid residues to the 3' end of the cDNA. The reaction product was extracted with phenol-chloroform and precipitated with ethanol to recover a cDNA-vector primer DNA to which a (dC) chain had been added.

The thus obtained (dC) chain-containing cDNA-vector primer DNA was dissolved in 400 µl of a solution (10 mM Tris-HCl, pH 7.5, 6 mM MgCl<sub>2</sub>, 60 mM NaCl), and 20 units of HindIII was also added. The mixture was incubated at 37°C for 2 hours to cut the DNA at HindIII site. The reaction product was extracted with phenol-chloroform and precipitated with ethanol. Thus, there was obtained 0.5 pmole of a (dC) chain-containing cDNA-vector primer DNA.

The thus obtained (dC) chain-containing cDNA-vector primer DNA (0.08 pmole) and the aforementioned linker DNA (0.16 pmole) were dissolved in 40 µl of a solution : 10 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 1 mM EDTA. The resulting solution was incubated at 65°C for 10 minutes, at 42°C for 25 minutes, and then at 0°C for 30 minutes. The reaction mixture was adjusted to 20 mM Tris-HCl (pH 7.5), 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M KCl and 0.1 mM β-NAD in a total volume of 400 µl.

To the reaction mixture, there was added 10 units of *E. coli* DNA ligase (New England Biolabs), followed by incubation overnight at 11°C. After adjusting the concentrations of dNTP and β-NAD in the reaction mixture to 40 µM and 0.15 mM, respectively, by supplementing necessary reagents, 5 units of *E. coli* DNA ligase, 7 units of *E. coli* DNA polymerase I (P-L Biochemicals) and 2 units of *E. coli* ribonuclease H (P-L Biochemicals) were added to the reaction mixture. The mixture was incubated at 12°C for one hour and then at 25°C for one hour.

In the course of the above reactions, a recombinant DNA containing the cDNA was cyclized and the RNA portion of the RNA-DNA double stranded chain was substituted by DNA. Thus, a desired recombinant plasmid containing a complete double-stranded DNA was produced.

The recombinant plasmid was used to transform competent cells of *E. coli* strain MC1064 prepared by conventional methods. Approximately 50,000 transformants were fixed on a nitrocellulose filter. These colonies were screened according to the colony hybridization method described in Molecular Cloning, Cold Spring Harbor Laboratory, p. 329 et seq. (1982) using the cDNA fragment obtained in Example 1 as a <sup>32</sup>P-labelled probe. Thus, three clones showed strong hybridization at 42°C.

These positive clones were analyzed in detail by Southern method : J. Mol. Biol., 98, 503 (1975). There was obtained the desired full length cDNA of a gene coding for an antigenic protein specific to non-A non-B hepatitis. The base sequence of the cDNA is shown in Fig. 4.

The expression vector containing the full length cDNA was designated as pCDVCL-I.

### EXAMPLE 3 : Preparation of Expression Vector and Transformant and Expression of Specific Antigen

#### A. Preparation of Expression Vector and Transformant

##### 1) Modification of N-terminus (Fig. 5) :

i) In 100 µl of a buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl<sub>2</sub>), pCDVCL-I (5 µg) was digested with PvuI (10 units) at 37°C for 2 hours. The reaction mixture was heated at 75°C for 15 minutes to deactivate the enzyme, dialyzed against water, and dried. The cleaved plasmid DNA was treated with T4 DNA polymerase (4 units) in 40 µl of a system : 33 mM Tris-acetic acid (pH 7.9), 66 mM potassium acetate, 10 mM magnesium acetate and 0.5 mM dithiothreitol, to which 2 mM 4-deoxytriphosphate had been added ; thus, the 3' protruding end of the plasmid DNA was filled in to produce a blunt end. The thus treated mixture was heated at 70°C for 10 minutes to deactivate the enzyme, dialyzed against water, and dried. The thus obtained plasmid DNA was then stored in the form of an aqueous solution (50 µl). This plasmid DNA fragment is hereinafter designated as Fragment I.

ii) On the other hand, pCDVCL-I (20 µg) was digested with NcoI and HindIII (each 20 units) at 37°C for 2 hours in 100 µl of a buffer : 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 6 mM MgCl<sub>2</sub>. The plasmid DNA was

subjected to 5% acrylamide gel electrophoresis at 10 V/cm for 1.5 hours in a buffer : 89 mM Tris, 89 mM boric acid, 2 mM EDTA. The gel was stained with 0.05% aqueous ethidium bromide solution and two gel slices corresponding to DNA fragments of larger molecular weights were excised from the gel under ultraviolet radiation at 340 nm. The gel slices were crushed by means of a glass rod, suspended into 4 ml of a buffer for DNA extraction (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% sodium laurylsulfate), and allowed to stand overnight at 37°C to extract DNA from the gel. The materials were subjected to centrifugation at 10,000 rpm for 15 minutes to eliminate larger gel pieces, and passed through a glass filter to remove smaller gel pieces. The DNA was purified by effecting ethanol precipitation three times and stored in the form of an aqueous solution (200 µl). This plasmid DNA fragment is hereinafter designated as Fragment II.

iii) A primer of the DNA portion to be modified as shown below (51 bases) was synthesized by a DNA synthesizer, NIKKAKI (Japan), Applied Biosystem MODEL 380A. The synthesized DNA was overnight reacted with concentrated aqueous ammonia at 55°C to deprotect and purified by reversed HPLC before use.

15  
 Primer HindIII  
ACAACAGATCTAAGCTTATGGCAGTTACAACAAGATTAA  
                  x  x  x                  x          xx  x  x  
 (Original sequence) (-----A--A-G-----G-----TC-C--G--  
 20  
CATGGTTGCATG wherein x represents a  
                  base substitution.  
-----)

25  
 The synthetic primer (150 pmole) was treated with T4 polynucleotide kinase (20 units) in 10 µl of a kinase buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl<sub>2</sub>, 5 mM dithiothreitol) to phosphorylate the 5' end thereof.  
 iv) Fragment I (0.05 pmole), Fragment II (0.05 pmole) and 5'-phosphorylated primer (45 pmole) were added to 12 µl of 5x polymerase-ligase buffer (0.5 M NaCl, 32.5 mM Tris-HCl, pH 7.5, 40 mM MgCl<sub>2</sub>, 5 mM β-mercaptoethanol) to make the total volume of the mixture 34.8 µl. The mixture was boiled at 100°C for 3 minutes, immediately after which it was placed in a thermostat at 30°C and allowed to stand for 30 minutes. The mixture was allowed to stand at 4°C for 30 minutes and then on ice for 10 minutes to form a heteroduplex.

30  
 To an aqueous solution (11.6 µl) containing the heteroduplex, there were added 2.5 mM 4-deoxynucleotide triphosphate (2 µl), 10 mM ATP (2 µl), Klenow enzyme (2 units) and T4 DNA ligase (0.5 units) to form a mixture of 20 µl in total volume. The mixture was reacted overnight at 16°C to cyclize the DNA.

35  
 An aqueous solution (2 µl) containing the circular DNA was used to transform *E. coli* HB101 strain according to conventional methods. Plasmids were separated from the transformant and purified in conventional manners. The plasmid was cleaved with restriction enzyme HindIII and subjected to 5% acrylamide gel electrophoresis.  
 40  
 Thus, two separate fragments were collected as desired modified, variant plasmids. Since resulting variant plasmids might often be admixed with original wild-type plasmids, the thus obtained variant plasmids were again employed to transform *E. coli* HB101 so as to purify the plasmid.

Thus, a purified plasmid pCV44H was obtained (Fig. 5).

#### 45 II) Modification of C-terminus (Fig. 6) :

i) Plasmid pCDVCL-I (5 µg) was treated in the same manner as in I) i) described above to produce Fragment I.

50 ii) Plasmid pCDVCL-I (20 µg) was treated in the same manner as in I) ii) described above except that *Nco*I and *Nsi*II (each 5 units) were employed. Thus, Fragment II was produced.

iii) In the same manner as in I) iii) described above, the following primer (46 bases) was synthesized and the 5' end thereof was phosphorylated.

Primer GCACAAGGAAAAAAT <sup>BglII</sup> <sup>SacI</sup> GAGATCTGTCGACGGTTCACGTA  
 (Original sequence) (-----AGATATGTGAA\*A-----)  
 AATTTCC wherein x represents a base  
 substitution and \* represents  
 -----) an addition.

iv) The Fragment I and II and the 5'-phosphorylated primer obtained above in ii) i) to iii) were treated in the same manner as in i) iv) described above. Thus, plasmid pCV44B was obtained (Fig. 6).

### iii) Introduction of cDNA coding for specific antigen into expression vector (Fig. 7):

i) In 100  $\mu$ l of a buffer H (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM  $MgCl_2$ ), 10  $\mu$ g (about 3 pmole) of pCV44H was cut with HindIII (20 units) and SacI (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis. Thus, a 467 bp DNA fragment coding for the N-terminus of the specific antigen was separated and purified. This fragment is hereinafter designated as Fragment N.

ii) In 100  $\mu$ l of the buffer H, 10  $\mu$ g (about 3 pmole) of pCV44B was cleaved with BglII (20 units) and SacI (20 units) at 37°C for 2 hours. The reaction mixture was subjected to 5% acrylamide gel electrophoresis to isolate and purify a 836 bp DNA fragment coding for the C-terminus of the specific antigen. The thus obtained fragment is hereinafter designated as Fragment C.

iii) In 20  $\mu$ l of buffer H, 2  $\mu$ g (about 1 pmole) of an expression vector pUS4H was cut with HindIII (2 units) and BglII (2 units) at 37°C for 2 hours. The reaction mixture was extracted with an equal volume of water-saturated phenol to remove proteins. After extracting the phenol with ether, the reaction mixture was dialyzed against water to desalt, and concentrated by a vacuum pump. Thus, there was obtained 10  $\mu$ l of an aqueous solution containing an expression vector fragment HB.

iv) Fragment N (0.5 pmole), Fragment C (0.5 pmole) and the expression vector fragment HB (0.1 pmole) were mixed and reacted with T4 DNA ligase (1 unit) at 4°C for 16 hours in 10  $\mu$ l of a buffer (10 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol, 6 mM  $MgCl_2$ , 1 mM ATP). The reaction mixture (3  $\mu$ l) was used to transform commercially available *E. coli* JM109 competent cell according to conventional methods. The resulting transformants were selected in L broth plate (bactopeptone 10 g, yeast extract 5 g, NaCl 10 g, agar 15 g per liter) containing 20  $\mu$ g/ml ampicillin. Thus, there was obtained an expression vector pCZ44 containing the specific antigen gene inserted thereinto (Fig. 7).

### B. Expression of Specific Antigen

*E. coli* strain JM109 possessing pCZ44 was cultured overnight at 30°C in L broth. The culture was inoculated in a fresh L broth with a dilution of 1/50 and cultured with shaking at 30°C for 2 hours. After IPTG (isopropylthio- $\beta$ -D-galactopyranoside) was added to the medium in a concentration of 2 mM, shaking culture was continued at 30°C for further 3 hours. The cells were collected by centrifugation at 6,500 rpm for 10 minutes and suspended in a buffer (0.9% NaCl, 10 mM Tris-HCl, pH 7.5) to store.

### C. Verification of Expression of Specific Antigen

The thus obtained cell culture (0.3 ml) was subjected to 10% SDS polyacrylamide gel electrophoresis at 120 V for one hour in a buffer (Tris 3g/l, glycine 14.4 g/l, 0.1% SDS). The gel was removed, placed on a nitrocellulose filter, interposed between filter papers and electrophoresed at 5 V/cm, 4°C in a buffer (Tris 3g/l, glycine 14.4 g/l) to transferred proteins in the gel onto the nitrocellulose filter. The nitrocellulose filter was rinsed with TBS buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl), immersed into 400 ml of TBS buffer containing 3% gelatine and shaken at 40°C for one hour to block the nitrocellulose filter.

To TBS buffer containing 1% gelatine, there was added a monoclonal antibody directed to a non-A non-B hepatitis-specific antigen ( $OD_{280} = 4.3$ ) with a dilution of 1/400. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 16 hours. The reaction mixture was washed three times with 400 ml of TBS buffer con-

aining 0.05% Tween 20 for 10 minutes.

To TBS buffer containing 1% gelatine, there was added a labelled secondary antibody, anti-mouse IgG-PAP (horseradish peroxidase, Bio Rad), with a dilution of 1/1000. The resulting mixture and the nitrocellulose filter were put into a vinyl bag so that the mixture was present in an amount of 2 ml per filter. Reaction was effected at room temperature for 2 hours. The reaction mixture was washed three times with 400 ml of TBS buffer containing 0.05% Tween 20 for 10 minutes.

Color formation was effected by immersing the filter into 20 ml of TBS buffer containing 12 mg of 4-chloro-1-naphthol (Bio Rad) and hydrogen peroxide. After completion of color formation, the filter was thoroughly washed with water, put into a vinyl bag containing water, and stored in a dark and cold place.

Such a test effected showed that a protein reacting with the monoclonal antibody was found at the same position (44 Kd) as found in the case of the specific antigen derived from infected chimpanzee liver. This verifies that such a specific antigen can be in fact expressed in *E coli*. The invention thus also relates to a process for the *in vitro* diagnosis of NON-A NON-B hepatitis, which comprises contacting a liver sample and/or a serum sample taken from a patient possibly infected with a NON-A NON-B hepatitis with the protein whose formula appears in claim 3 hereafter or a part thereof for a time and under conditions sufficient to allow for the production of a complex between said protein or part thereof with the antibodies contained in the patient sample and detecting the presence of the immunologic complex, particularly when the patient is suffering from NON-A NON-B hepatitis.

Any part of said protein, or any recombinant, produced by genetic engineering and including the amino acid sequence of said protein or part of said protein can be substituted for above-said protein, it being understood that the said recombinant protein or part of said protein are specifically recognized by the same antibodies as those which recognize said protein.

In other words the invention relates to all recombinant proteins or protein fragments which bind to antibodies contained in a liver extract or serum sample, or both, and originating from a patient suffering from NON-A NON-B hepatitis.

The invention also relates to a process for detecting *in vitro* an infection by a NON-A NON-B hepatitis virus, which process comprises contacting the DNA of claim 4, or a fragment thereof, under suitable hybridization conditions, with a sample of liver extract and/or serum sample originating from the patient to be diagnosed and in which the nucleic acid components had previously been made accessible to hybridization, to form a hybridization product between said DNA of claim 4 (probe) and the viral DNA of a NON-A NON-B hepatitis B virus, and detecting said hybridization product, particularly in the case where the patient is indeed infected with a NON-A NON-B virus.

### Claims

1. A DNA fragment which contains a base sequence coding for an antigenic protein specifically occurring in a host affected with non-A non-B hepatitis, said protein comprising the whole or a part of the amino acid sequence represented by the formula :

5 Met Ala Val Thr Thr Arg Leu Thr Trp 10 Leu His Glu Lys Ile Leu  
 Gln Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly 20 30  
 10 Ser Val His Gly Phe His Asn Gly Val 40 Leu Leu Asp Arg Cys Cys  
 Asn Gln Gly Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile 50 60  
 15 Ile Gly Ala Tyr Ala Glu Glu Gly Tyr Gln Glu Arg Lys Tyr Ala 70  
 Ser Ile Ile Leu Phe Ala Leu Gln Glu Thr Lys Ile Ser Glu Trp 80 90  
 20 Lys Leu Gly Leu Tyr Thr Pro Glu Thr 100 Leu Phe Cys Cys Asp Val  
 Ala Lys Tyr Asn Ser Pro Thr Asn Phe Gln Ile Asp Gly Arg Asn 110 120  
 25 Arg Lys Val Ile Met Asp Leu Lys Thr 130 Met Glu Asn Leu Gly Leu  
 Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg 140 150  
 Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu 160  
 35 Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr Tyr Glu Pro Tyr 170 180  
 40 Gly Ser Leu Val Gln Gln Ile Arg Ile 190 Leu Leu Leu Gly Pro Ile

45

50

55

200 210  
 Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe  
 5  
 220  
 Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr  
 230 240  
 10 Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys  
 250  
 Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu  
 15 260 270  
 Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile  
 280  
 Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu  
 20 290 300  
 Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu  
 310  
 25 Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser  
 320 330  
 Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile  
 340  
 30 Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu  
 350 360  
 Thr His Val Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile  
 35 370  
 Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val  
 380 390  
 40 Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser  
 400  
 Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu  
 45 410 420  
 Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe  
 430  
 Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu  
 50 440  
 Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*.

55

2. The DNA fragment in accordance with claim 1, in which the base sequence comprises the whole or a part of the base sequence represented by the formula :

5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG CAT GAA AAG ATC  
 3' --- --- --- --- --- --- --- --- --- --- --- --- ---  
 10  
 CTG CAA AAT CAT TTT GGA GGG AAG CGG CTT AGC CTT CTC TAT  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 15  
 AAG GGT AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG CTT GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 18  
 AGA TGT TGT AAT CAA GGG CCT ACT CTA ACA GTG ATT TAT AGT  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 20  
 GAA GAT CAT ATT ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 25  
 GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT GCA CTT CAA GAG  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 30  
 ACT AAA ATT TCA GAA TGG AAA CTA GGA CTA TAT ACA CCA GAA  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 35  
 ACA CTG TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA ACT  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 40  
 45  
 50  
 55



	340	350	360	370	
	AAT TTC CAG	ATA GAT GGA	AGA AAT AGA	AAA GTG ATT	ATG GAC
5	---	---	---	---	---
	380	390	400	410	420
	TTA AAG ACA	ATG GAA AAT	CTT GGA CTT	GCT CAA AAT	TGT ACT
	---	---	---	---	---
10		430	440	450	460
	ATC TCT ATT	CAG GAT TAT	GAA GTT TTT	CGA TGC GAA	GAT TCA
	---	---	---	---	---
15	470	480	490	500	
	CTG GAC GAA	AGA AAG ATA	AAA GGG GTC	ATT GAG CTC	AGG AAG
	---	---	---	---	---
20	510	520	530	540	
	AGC TTA CTG	TCT GCC TTG	AGA ACT TAT	GAA CCA TAT	GGA TCC
	---	---	---	---	---
25	550	560	570	580	
	CTG GTT CAA	CAA ATA CGA	ATT CTG CTG	CTG GGT CCA	ATT GGA
	---	---	---	---	---
30	590	600	610	620	630
	GCT GGG AAG	TCT AGC TTT	TTC AAC TCA	GTG AGG TCT	GTT TTC
	---	---	---	---	---
35	640	650	660	670	
	CAA GGG CAT	GTA ACG CAT	CAG GCT TTG	GTG GGC ACT	AAT ACA
	---	---	---	---	---
40	680	690	700	710	
	ACT GGG ATA	TCT GAG AAG	TAT AGG ACA	TAC TCT ATT	AGA GAC
	---	---	---	---	---
45	720	730	740	750	
	GGG AAA GAT	GGC AAA TAC	CTG CCA TTT	ATT CTG TGT	GAC TCA
	---	---	---	---	---
50	760	770	780	790	
	CTG GGG CTG	AGT GAG AAA	GAA GGC GGC	CTG TGC ATG	GAT GAC
	---	---	---	---	---
55	800	810	820	830	840
	ATA TCC TAC	ATC TTG AAC	GGT AAC ATT	CGT GAT AGA	TAC CAG
	---	---	---	---	---
60	850	860	870	880	
	TTT AAT CCC	ATG GAA TCA	ATC AAA TTA	AAT CAT CAT	GAC TAC
	---	---	---	---	---

890 900 910 920  
 ATT GAT TCC CCA TCG CTG AAG GAC AGA ATT CAT TGT GTG GCA  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 5  
 930 940 950 960  
 TTT GTA TTT GAT GCC AGC TCT ATT GAA TAC TTC TCC TCT CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 10  
 970 980 990 1000  
 ATG ATA GTA AAG ATC AAA AGA ATT CGA AGG GAG TTG GTA AAC  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 15  
 1010 1020 1030 1040 1050  
 GCT GGT GTG GTA CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 20  
 1060 1070 1080 1090  
 ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 25  
 1100 1110 1120 1130  
 TGT GTG CCT GTG AGG TCC AAG CTA GAG GAA GTC CAA AGA AAA  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 30  
 1140 1150 1160 1170  
 CTT GGA TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT TAT  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 35  
 1180 1190 1200 1210  
 TCC TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GTT CTA ATT  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 40  
 1220 1230 1240 1250 1260  
 CTT TCT GCT CTG AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 45  
 1270 1280 1290 1300  
 TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG AAT CTA AGG GAG  
 --- --- --- --- --- --- --- --- --- --- --- --- ---  
 50  
 1310 1320 1330  
 GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA 3'  
 --- --- --- --- --- --- --- --- --- 5'

wherein the sign "-" represents a base complementary to the base shown just above each sign.

3. An expression vector in which a DNA fragment containing a base sequence according to claim 1 or 2 and coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis is introduced into a cloning site present downstream from a promoter of said vector.

4. The expression vector in accordance with claim 3, in which the promoter is controllable by a regulatory factor.

5. The expression vector in accordance with claim 3, in which the promoter operates in a microorganism.

6. The expression vector in accordance with claim 3, in which the promoter operates in an eukaryote.

55 7. A transformant obtained by transforming a host with an expression vector in which a DNA fragment containing a base sequence according to claim 1 or 2 and coding for an antigen specifically occurring in a host affected with non-A non-B hepatitis is introduced into a cloning site present downstream from a promoter of said vector.

8. The transformant in accordance to claim 7, in which the host is Escherichia coli or Bacillus subtilis.

9. A process for producing an antigen occurring specifically in a host affected with non-A non-B hepatitis, comprising introducing a DNA fragment containing a base sequence according to claim 1 or 2 and coding for said specifically occurring antigen into a cloning site present downstream from a promoter of a vector for expression, introducing the expression vector containing said DNA fragment into a host, culturing said transformed, and collecting the produced and accumulated antigen.

10. Process for the in vitro diagnosis of non-A non-B hepatitis which comprises contacting a liver sample and/or a serum sample taken from a patient suspected of being infected with a non-A non-B hepatitis, with the protein whose sequence appears in claim 1, or a part thereof, for a time and under conditions sufficient to allow for the production of a complex between said protein or a part thereof with the antibodies contained in the patient sample and detecting the presence of the immunological complex, particularly when the patient is suffering from non-A non-B hepatitis.

## 15 Revendications

1. Fragment d'ADN qui contient une séquence de bases codant pour une protéine antigénique rencontrée spécifiquement chez un hôte atteint de l'hépatite non-A non-B, ladite protéine comprenant la totalité ou une partie de la séquence d'acides aminés représentée par la formule :

10  
 Met Ala Val Thr Thr Arg Leu Thr Trp Leu His Glu Lys Ile Leu  
 5  
 20  
 Gln Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly  
 30  
 40  
 Ser Val His Gly Phe His Asn Gly Val Leu Leu Asp Arg Cys Cys  
 10  
 50  
 Asn Gln Gly Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile  
 60  
 70  
 Ile Gly Ala Tyr Ala Glu Glu Gly Tyr Gln Glu Arg Lys Tyr Ala  
 15  
 80  
 Ser Ile Ile Leu Phe Ala Leu Gln Glu Thr Lys Ile Ser Glu Trp  
 20  
 90  
 Lys Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val  
 100  
 110  
 Ala Lys Tyr Asn Ser Pro Thr Asn Phe Gln Ile Asp Gly Arg Asn  
 25  
 120  
 Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu Asn Leu Gly Leu  
 30  
 130  
 Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg  
 140  
 150  
 Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu  
 35  
 160  
 Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr Tyr Glu Pro Tyr  
 170  
 180  
 Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu Leu Gly Pro Ile  
 40  
 190  
 45  
 50  
 55

200 210  
 5 Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe  
 220  
 Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr  
 230 240  
 10 Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys  
 250  
 Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu  
 15 260 270  
 Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile  
 280  
 20 Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu  
 290 300  
 Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu  
 310  
 25 Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser  
 320 330  
 30 Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile  
 340  
 Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu  
 350 360  
 35 Thr His Val Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile  
 370  
 Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val  
 40 380 390  
 Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser  
 400  
 45 Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu  
 410 420  
 Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe  
 430  
 50 Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu  
 440  
 Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*.

55

2. Fragment d'ADN selon la revendication 1, dans lequel la séquence de bases comprend la totalité ou une partie de la séquence de bases représentée par la formule :

5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG CAT GAA AAG ATC  
 3' --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 5  
 CTG CAA AAT CAT TTT GGA GGG AAG CGG CTT AGC CTT CTC TAT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 10  
 AAG GGT AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG CTT GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 15  
 AGA TGT TGT AAT CAA GGG CCT ACT CTA ACA GTG ATT TAT AGT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 20  
 GAA GAT CAT ATT ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 25  
 GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT GCA CTT CAA GAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 30  
 ACT AAA ATT TCA GAA TGG AAA CTA GGA CTA TAT ACA CCA GAA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 35  
 ACA CTG TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA ACT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 40  
 45  
 50  
 55

340                      350                      360                      370  
 AAT TTC CAG ATA GAT GGA AGA AAT AGA AAA GTG ATT ATG GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 5  
 380                      390                      400                      410                      420  
 TTA AAG ACA ATG GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 10  
 430                      440                      450                      460  
 ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC GAA GAT TCA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 15  
 470                      480                      490                      500  
 CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 20  
 510                      520                      530                      540  
 AGC TTA CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 25  
 550                      560                      570                      580  
 CTG GTT CAA CAA ATA CGA ATT CTG CTG CTG GGT CCA ATT GGA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 30  
 590                      600                      610                      620                      630  
 GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 35  
 640                      650                      660                      670  
 CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 40  
 680                      690                      700                      710  
 ACT GGG ATA TCT GAG AAG TAT AGG ACA TAC TCT ATT AGA GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 45  
 720                      730                      740                      750  
 GGG AAA GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT GAC TCA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 50  
 760                      770                      780                      790  
 CTG GGG CTG AGT GAG AAA GAA GGC GGC CTG TGC ATG GAT GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 55  
 800                      810                      820                      830                      840  
 ATA TCC TAC ATC TTG AAC GGT AAC ATT CGT GAT AGA TAC CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 850                      860                      870                      880  
 TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT CAT CAT GAC TAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---

```

      890      900      910      920
ATT GAT TCC CCA TCG CTG AAG GAC AGA ATT CAT TGT GTG GCA
---
5
      930      940      950      960
TTT GTA TTT GAT GCC AGC TCT ATT GAA TAC TTC TCC TCT CAG
---
10
      970      980      990      1000
ATG ATA GTA AAG ATC AAA AGA ATT CGA AGG GAG TTG GTA AAC
---
15
1010      1020      1030      1040      1050
GCT GGT GTG GTA CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC
---
      1060      1070      1080      1090
20
ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA
---
      1100      1110      1120      1130
25
TGT GTG CCT GTG AGG TCC AAG CTA GAG GAA GTC CAA AGA AAA
---
      1140      1150      1160      1170
      CTT GGA TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT TAT
30
---
      1180      1190      1200      1210
TCC TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GTT CTA ATT
---
35
1220      1230      1240      1250      1260
CTT TCT GCT CTG AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC
---
      1270      1280      1290      1300
40
TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG AAT CTA AGG GAG
---
      1310      1320      1330
45
GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA 3'
---
5'

```

dans laquelle le signe "-" représente une base complémentaire de la base représentée juste au-dessus de chaque signe.

50 3. Vecteur d'expression dans lequel un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour un antigène rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, est introduit dans un site de clonage présent en aval d'un promoteur dudit vecteur.

4. Vecteur d'expression selon la revendication 3, dans lequel le promoteur peut être contrôlé par un facteur de régulation.

55 5. Vecteur d'expression selon la revendication 3, dans lequel le promoteur opère dans un microorganisme.

6. Vecteur d'expression selon la revendication 3, dans lequel le promoteur opère dans un eucaryote.

7. Transformant obtenu par transformation d'un hôte par un vecteur d'expression dans lequel un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour un antigène



rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, est introduit dans un site de clonage présent en aval d'un promoteur dudit vecteur.

8. Transformant selon la revendication 7, dans lequel l'hôte est *Escherichia coli* ou *Bacillus subtilis*.

5 9. Procédé de production d'un antigène rencontré spécifiquement chez un hôte atteint de l'hépatite non-A non-B, comprenant l'introduction d'un fragment d'ADN contenant une séquence de bases telle que définie à la revendication 1 ou 2 et codant pour ledit antigène rencontré spécifiquement dans un site de clonage présent en aval d'un promoteur d'un vecteur d'expression, l'introduction du vecteur d'expression contenant ledit fragment d'ADN dans un hôte, la culture dudit transformant, et la récupération de l'antigène produit et accumulé.

10 10. Procédé pour le diagnostic *in vitro* de l'hépatite non-A non-B, qui comprend la mise en contact d'un échantillon de foie et/ou d'un échantillon de sérum prélevé chez un patient soupçonné être infecté par une hépatite non-A non-B, avec la protéine dont la séquence apparaît à la revendication 1, ou une partie de celle-ci, pendant un laps de temps suffisant et dans des conditions suffisantes pour permettre la production d'un complexe entre ladite protéine ou une partie de celle-ci, avec les anticorps contenus dans l'échantillon du patient, et la détection de la présence du complexe immunologique, en particulier, lorsque le patient souffre  
15 d'une hépatite non-A non-B.

#### Patentansprüche

20 1. DNA-Fragment, das eine für ein antigenes Protein, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt vorkommt, kodierende Basensequenz enthält, wobei das Protein die ganze oder einen Teil der durch die Formel dargestellten Aminosäuresequenz enthält :

25

30

35

40

45

50

55

10  
 Met Ala Val Thr Thr Arg Leu Thr Trp Leu His Glu Lys Ile Leu  
 5  
 20  
 Gln Asn His Phe Gly Gly Lys Arg Leu Ser Leu Leu Tyr Lys Gly  
 30  
 40  
 10 Ser Val His Gly Phe His Asn Gly Val Leu Leu Asp Arg Cys Cys  
 50  
 Asn Gln Gly Pro Thr Leu Thr Val Ile Tyr Ser Glu Asp His Ile  
 60  
 70  
 15 Ile Gly Ala Tyr Ala Glu Glu Gly Tyr Gln Glu Arg Lys Tyr Ala  
 80  
 20 Ser Ile Ile Leu Phe Ala Leu Gln Glu Thr Lys Ile Ser Glu Trp  
 90  
 100  
 Lys Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val  
 110  
 25 Ala Lys Tyr Asn Ser Pro Thr Asn Phe Gln Ile Asp Gly Arg Asn  
 120  
 130  
 Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu Asn Leu Gly Leu  
 30  
 140  
 Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg  
 150  
 160  
 35 Cys Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu  
 170  
 Leu Arg Lys Ser Leu Leu Ser Ala Leu Arg Thr Tyr Glu Pro Tyr  
 180  
 40 190  
 Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu Leu Gly Pro Ile  
 190

200 210  
 (Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe  
 5 220  
 (Gln Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr  
 230 240  
 10 Gly Ile Ser Glu Lys Tyr Arg Thr Tyr Ser Ile Arg Asp Gly Lys  
 250  
 Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp Ser Leu Gly Leu  
 15 260 270  
 Ser Glu Lys Glu Gly Gly Leu Cys Met Asp Asp Ile Ser Tyr Ile  
 280  
 20 Leu Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu  
 290 300  
 Ser Ile Lys Leu Asn His His Asp Tyr Ile Asp Ser Pro Ser Leu  
 310  
 25 Lys Asp Arg Ile His Cys Val Ala Phe Val Phe Asp Ala Ser Ser  
 320 330  
 Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile  
 30 340  
 Arg Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu  
 350 360  
 35 Thr His Val Asp Ser Met Asp Leu Ile Thr Lys Gly Asp Leu Ile  
 370  
 Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys Leu Glu Glu Val  
 380 390  
 40 Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser  
 400  
 45 Asn Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu  
 410 420  
 Ile Leu Ser Ala Leu Arg Arg Met Leu Trp Ala Ala Asp Asp Phe  
 430  
 50 Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn Leu Arg Glu Glu  
 440  
 Ile Ile Asn Cys Ala Gln Gly Lys Lys \*\*\*.

55

2. DNA-Fragment nach Anspruch 1, bei dem die Basensequenz die ganze oder einen Teil der durch die Formel dargestellten Phasensequenz enthält:

5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG CAT GAA AAG ATC  
 3' --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 5  
 CTG CAA AAT CAT TTT GGA GGG AAG CGG CTT AGC CTT CTC TAT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 10  
 AAG GGT AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG CTT GAC  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 15  
 AGA TGT TGT AAT CAA GGG CCT ACT CTA ACA GTG ATT TAT AGT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 20  
 GAA GAT CAT ATT ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 25  
 GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT GCA CTT CAA GAG  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 30  
 ACT AAA ATT TCA GAA TGG AAA CTA GGA CTA TAT ACA CCA GAA  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 35  
 ACA CTG TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA ACT  
 --- --- --- --- --- --- --- --- --- --- --- --- --- ---  
 40  
 45  
 50  
 55

	340	350	360	370	
	AAT TTC CAG	ATA GAT GGA	AGA AAT AGA	AAA GTG ATT	ATG GAC
	---	---	---	---	---
5	380	390	400	410	420
	TTA AAG ACA	ATG GAA AAT	CTT GGA CTT	GCT CAA AAT	TGT ACT
	---	---	---	---	---
10		430	440	450	460
	ATC TCT ATT	CAG GAT TAT	GAA GTT TTT	CGA TGC GAA	GAT TCA
	---	---	---	---	---
15		470	480	490	500
	CTG GAC GAA	AGA AAG ATA	AAA GGG GTC	ATT GAG CTC	AGG AAG
	---	---	---	---	---
20	510	520	530	540	
	AGC TTA CTG	TCT GCC TTG	AGA ACT TAT	GAA CCA TAT	GGA TCC
	---	---	---	---	---
25	550	560	570	580	
	CTG GTT CAA	CAA ATA CGA	ATT CTG CTG	CTG GGT CCA	ATT GGA
	---	---	---	---	---
30	590	600	610	620	630
	GCT GGG AAG	TCT AGC TTT	TTC AAC TCA	GTG AGG TCT	GTT TTC
	---	---	---	---	---
35		640	650	660	670
	CAA GGG CAT	GTA ACG CAT	CAG GCT TTG	GTG GGC ACT	AAT ACA
	---	---	---	---	---
40		680	690	700	710
	ACT GGG ATA	TCT GAG AAG	TAT AGG ACA	TAC TCT ATT	AGA GAC
	---	---	---	---	---
45		720	730	740	750
	GGG AAA GAT	GGC AAA TAC	CTG CCA TTT	ATT CTG TGT	GAC TCA
	---	---	---	---	---
50		760	770	780	790
	CTG GGG CTG	AGT GAG AAA	GAA GGC GGC	CTG TGC ATG	GAT GAC
	---	---	---	---	---
55	800	810	820	830	840
	ATA TCC TAC	ATC TTG AAC	GGT AAC ATT	CGT GAT AGA	TAC CAG
	---	---	---	---	---
		850	860	870	880
	TTT AAT CCC	ATG GAA TCA	ATC AAA TTA	AAT CAT CAT	GAC TAC
	---	---	---	---	---

```

      890      900      910      920
ATT GAT TCC CCA TCG CTG AAG GAC AGA ATT CAT TGT GTG GCA
5  --- --- --- --- --- --- --- --- --- --- --- --- ---

      930      940      950      960
TTT GTA TTT GAT GCC AGC TCT ATT GAA TAC TTC TCC TCT CAG
10 --- --- --- --- --- --- --- --- --- --- --- --- ---

      970      980      990      1000
ATG ATA GTA AAG ATC AAA AGA ATT CGA AGG GAG TTG GTA AAC
15 --- --- --- --- --- --- --- --- --- --- --- --- ---

1010      1020      1030      1040      1050
GCT GGT GTG GTA CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC
--- --- --- --- --- --- --- --- --- --- --- --- ---

      1060      1070      1080      1090
ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA
20 --- --- --- --- --- --- --- --- --- --- --- --- ---

      1100      1110      1120      1130
TGT GTG CCT GTG AGG TCC AAG CTA GAG GAA GTC CAA AGA AAA
25 --- --- --- --- --- --- --- --- --- --- --- --- ---

      1140      1150      1160      1170
CTT GGA TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT TAT
30 --- --- --- --- --- --- --- --- --- --- --- --- ---

      1180      1190      1200      1210
TCC TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GTT CTA ATT
35 --- --- --- --- --- --- --- --- --- --- --- --- ---

1220      1230      1240      1250      1260
CTT TCT GCT CTG AGA CGA ATG CTA TGG GCT GCA GAT GAC TTC
--- --- --- --- --- --- --- --- --- --- --- --- ---

40      1270      1280      1290      1300
TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG AAT CTA AGG GAG
--- --- --- --- --- --- --- --- --- --- --- --- ---

45      1310      1320      1330
GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA 3'
--- --- --- --- --- --- --- --- --- 5'

```

wobei "-" jeweils die zur direkt darüberstehenden Base komplementäre Base darstellt.

50 3. Expressionsvektor, bei dem ein DNA-Fragment, das eine Basensequenz nach Anspruch 1 oder 2 enthält und für ein Antigen kodiert, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt auftritt, in eine stromabwärts vom Promotor des Vektors vorhandene Klonierungsstelle eingeführt wird.

4. Expressionsvektor nach Anspruch 3, bei dem der Promotor durch einen Regulationsfaktor kontrollierbar ist.

55 5. Expressionsvektor nach Anspruch 3, bei dem der Promotor in einem Mikroorganismus wirksam ist.

6. Expressionsvektor nach Anspruch 3, bei dem der Promotor in einem Eukaryonten wirksam ist.

7. Transformante, erhalten durch Transformieren eines Wirtes mit einem Expressionsvektor, bei dem ein DNA-Fragment, das eine Basensequenz nach Anspruch 1 oder 2 enthält, die für ein Antigen kodiert, das spe-

zifisch in einem mit non-A non-B Hepatitis befallenen Wirt auftritt, in eine stromabwärts vom Promotor des Vektors vorhandene Klonierungsstelle eingeführt wird.

8. Transformante nach Anspruch 7, bei der der Wirt Escherichia coli oder Bacillus subtilis ist.

5 9. Verfahren zur Herstellung eines Antigens, das spezifisch in einem mit non-A non-B Hepatitis befallenen Wirt vorkommt, umfassend die Schritte : Einführen eines DNA-Fragments, das eine Basensequenz nach Anspruch 1 oder 2 enthält, die für das spezifisch vorkommende Antigen kodiert, in eine stromabwärts vom Promotor eines Expressionsvektors vorhandene Klonierungsstelle, Einführen des das DNA-Fragment enthalten- den Expressionsvektors in einen Wirt, Kultivieren der Transformanten und Gewinnung des produzierten und akkumulierten Antigens.

10 10. Verfahren zur in vitro Diagnose von non-A non-B Hepatitis, umfassend die Schritte : in Kontakt bringen einer Leberprobe und/oder einer Serumprobe, die von einem Patienten mit Verdacht auf eine non-A non-B Hepatitisinfektion entnommen wurde, mit dem Protein mit einer Sequenz wie in Anspruch 1, oder einem Teil davon, für eine bestimmte Zeit und unter Bedingungen, die ausreichend sind für die Komplexbildung zwischen dem Protein, oder einem Teil davon, mit dem in der Patientenprobe enthaltenen Antikörpern, und Nachweis  
15 der Anwesenheit des immunologischen Komplexes, insbesondere wenn der Patient an non-A non-B Hepatitis leidet.

20

25

30

35

40

45

50

55

Fig. 1a

```

      10      20      30
5' ATG GCA GTG ACA ACT CGT TTG ACA TGG TTG
3'

      40      50      60
   CAT GAA AAG ATC CTG CAA AAT CAT TTT GGA

      70      80      90
   GGG AAG CGG CTT AGC CTT CTC TAT AAG GGT

     100     110     120
   AGT GTC CAT GGA TTC CAT AAT GGA GTT TTG

     130     140     150
   CTT GAC AGA TGT TGT AAT CAA GGG CCT ACT

     160     170     180
   CTA ACA GTG ATT TAT AGT GAA GAT CAT ATT

     190     200     210
   ATT GGA GCA TAT GCA GAA GAG GGT TAC CAG

     220     230     240
   GAA AGA AAG TAT GCT TCC ATC ATC CTT TTT

     250     260     270
   GCA CTT CAA GAG ACT AAA ATT TCA GAA TGG

     280     290     300
   AAA CTA GGA CTA TAT ACA CCA GAA ACA CTG

```



Fig. 1b

310                      320                      330  
TTT TGT TGT GAC GTT GCA AAA TAT AAC TCC  
  
 340                      350                      360  
CCA ACT AAT TTC CAG ATA GAT GGA AGA AAT  
  
 370                      380                      390  
AGA AAA GTG ATT ATG GAC TTA AAG ACA ATG  
  
 400                      410                      420  
GAA AAT CTT GGA CTT GCT CAA AAT TGT ACT  
  
 430                      440                      450  
ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA  
  
 460                      470                      480  
TGC GAA GAT TCA CTG GAC GAA AGA AAG ATA  
  
 490                      500                      510  
AAA GGG GTC ATT GAG CTC AGG AAG AGC TTA  
  
 520                      530                      540  
CTG TCT GCC TTG AGA ACT TAT GAA CCA TAT  
  
 550                      560                      570  
GGA TCC CTG GTT CAA CAA ATA CGA ATT CTG  
  
 580                      590                      600  
CTG CTG GGT CCA ATT GGA GCT GGG AAG TCT

Fig. 1c

610 620 630  
AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC

640 650 660  
CAA GGG CAT GTA ACG CAT CAG GCT TTG GTG

670 680 690  
GGC ACT AAT ACA ACT GGG ATA TCT GAG AAG

700 710 720  
TAT AGG ACA TAC TCT ATT AGA GAC GGG AAA

730 740 750  
GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT

760 770 780  
GAC TCA CTG GGG CTG AGT GAG AAA GAA GGC

790 800 810  
GGC CTG TGC ATG GAT GAC ATA TCC TAC ATC

820 830 840  
TTG AAC GGT AAC ATT CGT GAT AGA TAC CAG

850 860 870  
TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT

880 890 900  
CAT CAT GAC TAC ATT GAT TCC CCA TCG CTG

Fig. 1d

910                      920                      930  
AAG GAC AGA ATT CAT TGT GTG GCA TTT GTA

940                      950                      960  
TTT GAT GCC AGC TCT ATT GAA TAC TTC TCC

970                      980                      990  
TCT CAG ATG ATA GTA AAG ATC AAA AGA ATT

1000                      1010                      1020  
CGA AGG GAG TTG GTA AAC GCT GGT GTG GTA

1030                      1040                      1050  
CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC

1060                      1070                      1080  
ATG GAT CTG ATT ACA AAA GGT GAC CTT ATA

1090                      1100                      1110  
GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC

1120                      1130                      1140  
AAG CTA GAG GAA GTC CAA AGA AAA CTT GGA

1150                      1160                      1170  
TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC

1180                      1190                      1200  
AAT TAT TCC TCT GAG TGG GAG CTG GAC CCT

Fig. 1e

	1210	1220	1230
<u>GTA</u>	<u>AAG</u>	<u>GAT</u>	<u>GTT</u>
<u>CTA</u>	<u>ATT</u>	<u>CTT</u>	<u>TCT</u>
<u>GCT</u>	<u>CTG</u>		
	1240	1250	1260
<u>AGA</u>	<u>CGA</u>	<u>ATG</u>	<u>CTA</u>
<u>TGG</u>	<u>GCT</u>	<u>GCA</u>	<u>GAT</u>
<u>GAC</u>	<u>TTC</u>		
	1270	1280	1290
<u>TTA</u>	<u>GAG</u>	<u>GAT</u>	<u>TTG</u>
<u>CCT</u>	<u>TTT</u>	<u>GAG</u>	<u>CAA</u>
<u>ATA</u>	<u>GGG</u>		
	1300	1310	1320
<u>AAT</u>	<u>CTA</u>	<u>AGG</u>	<u>GAG</u>
<u>GAA</u>	<u>ATT</u>	<u>ATC</u>	<u>AAC</u>
<u>TGT</u>	<u>GCA</u>		
	1330		
<u>CAA</u>	<u>GGA</u>	<u>AAA</u>	<u>AAA</u>
		3'	
		5'	

Fig. 2

AAAAATTATTGCTTTCAGGAAAATTTTCTGT  
TTTTTAAATAAACGAAAGTCCTTTTAAAAAGACA

ATAATGTGTGGAATTGTGAGCGGATAACAATTTC  
TATTACACACCTTAACACTCGCCTATTGTAAAG

Fig. 3a

250 CTA GGA CTA TAT ACA CCA GAA ACA CTG TTT TGT TGT TGT GAC GTT GCA AAA TAT AAC TCC CCA 300  
 Leu Gly Leu Tyr Thr Pro Glu Thr Leu Phe Cys Cys Asp Val Ala Lys Tyr Asn Ser Pro

310 ACT AAT TTC CAG ATA GAT GGA AGA AAT AGA AAA GTG ATT ATG GAC TTA AAG ACA ATG GAA 360  
 Thr Asn Phe Gln Ile Asp Gly Arg Asn Arg Lys Val Ile Met Asp Leu Lys Thr Met Glu

370 AAT CTT GGA CTT GCT CAA AAT TGT ACT ATC TCT ATT CAG GAT TAT GAA GTT TTT CGA TGC 420  
 Asn Leu Gly Leu Ala Gln Asn Cys Thr Ile Ser Ile Gln Asp Tyr Glu Val Phe Arg Cys

430 GAA GAT TCA CTG GAC GAA AGA AAG ATA AAA GGG GTC ATT GAG CTC AGG AAG AGC TTA CTG 480  
 Glu Asp Ser Leu Asp Glu Arg Lys Ile Lys Gly Val Ile Glu Leu Arg Lys Ser Leu Leu

490 TCT GCC TTG AGA ACT TAT GAA CCA TAT GGA TCC CTG GTT CAA CAA ATA CGA ATT CTG CTG 540  
 Ser Ala Leu Arg Thr Tyr Glu Pro Tyr Gly Ser Leu Val Gln Gln Ile Arg Ile Leu Leu

550 CTG GGT CCA ATT GGA GCT GGG AAG TCT AGC TTT TTC AAC TCA GTG AGG TCT GTT TTC CAA 600  
 Leu Gly Pro Ile Gly Ala Gly Lys Ser Ser Phe Phe Asn Ser Val Arg Ser Val Phe Gln

Fig. 3b

610 GGG CAT GTA ACG CAT CAG GCT TTG GTG GGC ACT AAT ACA ACT GGG ATA TCT GAG AAG TAT 660  
 Gly His Val Thr His Gln Ala Leu Val Gly Thr Asn Thr Thr Gly Ile Ser Glu Lys Tyr  
 670 AGG ACA TAC TCT ATT AGA GAC GGC AAA GAT GGC AAA TAC CTG CCA TTT ATT CTG TGT GAC 720  
 Arg Thr Tyr Ser Ile Arg Asp Gly Lys Asp Gly Lys Tyr Leu Pro Phe Ile Leu Cys Asp  
 730 TCA CTG GGC CTG AGT GAG AAA GAA GGC GGC CTG TGC ATG GAT GAC ATA TCC TAC ATC TTG 780  
 Ser Leu Gly Leu Ser Glu Lys Glu Gly Leu Cys Met Asp Ile Ser Tyr Ile Leu  
 790 AAC GGT AAC ATT CGT GAT AGA TAC CAG TTT AAT CCC ATG GAA TCA ATC AAA TTA AAT CAT 840  
 Asn Gly Asn Ile Arg Asp Arg Tyr Gln Phe Asn Pro Met Glu Ser Ile Lys Leu Asn His  
 850 CAT GAC TAC ATT GAT TCC CCA TCG CTG AAG GAC AGA ATT CAT TGT GTG GCA TTT GTA TTT 900  
 His Asp Tyr Ile Asp Ser Pro Ser Leu Lys Asp Arg Ile His Cys Val Ala Phe Val Phe  
 910 GAT GCC AGC TCT ATT GAA TAC TTC TCC TCT CAG ATG ATA GTA AAG ATC AAA AGA ATT CGA 960  
 Asp Ala Ser Ser Ile Glu Tyr Phe Ser Ser Gln Met Ile Val Lys Ile Lys Arg Ile Arg  
 970 AGG GAG TTG GTA AAC GCT GGT GTG GTA CAT GTG GCT TTG CTC ACT CAT GTG GAT AGC ATG 1020  
 Arg Glu Leu Val Asn Ala Gly Val Val His Val Ala Leu Leu Thr His Val Asp Ser Met

Fig. 3c

1030	1040	1050	1060	1070	1080
GAT CTG ATT ACA AAA GGT GAC CTT ATA GAA ATA GAG AGA TGT GTG CCT GTG AGG TCC AAG					
Asp Leu Ile Thr Lys Lys Gly Asp Leu Ile Glu Ile Glu Arg Cys Val Pro Val Arg Ser Lys					
1090	1100	1110	1120	1130	1140
CTA GAG GAA GTC CAA AGA AAA CTT GGA TTT GCT CTT TCT GAC ATC TCG GTG GTT AGC AAT					
Leu Glu Glu Val Gln Arg Lys Leu Gly Phe Ala Leu Ser Asp Ile Ser Val Val Ser Asn					
1150	1160	1170	1180	1190	1200
TAT TCC TCT GAG TGG GAG CTG GAC CCT GTA AAG GAT GTT CTA ATT CTT TCT GCT CTG AGA					
Tyr Ser Ser Glu Trp Glu Leu Asp Pro Val Lys Asp Val Leu Ile Leu Ser Ala Leu Arg					
1210	1220	1230	1240	1250	1260
CGA ATG CTA TGG GCT GCA GAT GAC TTC TTA GAG GAT TTG CCT TTT GAG CAA ATA GGG AAT					
Arg Met Leu Trp Ala Ala Asp Asp Phe Leu Glu Asp Leu Pro Phe Glu Gln Ile Gly Asn					
1270	1280	1290	1300		
CTA AGG GAG GAA ATT ATC AAC TGT GCA CAA GGA AAA AAA TAG					
Leu Arg Glu Glu Ile Ile Asn Cys Ala Gln Gly Lys Lys ***					



Fig. 4a

10 20 30 40 50 60 70 80  
 5' GGGGGGCTAC CCTCAGCTCT AGCTCATACT ACAGACAGTA CAACAGATCA AGAAGTATGG CAGTGACAAC TCGTTTGACA  
 3' CCCCCCGATG GGAGTCGAGA TCGAGTATGA TGTCTGTCAAT TGTCTGTAGT TCTTCATACC GTCACCTGTTG AGCAAACTGT

90 100 110 120 130 140 150 160  
 TGGTTGCATG AAAAGATCCT GCAAAATCAT TTTGGAGGGA AGCGGCTTAG CCTTCTCTAT AAGGGTAGTG TCCATGGATT  
 ACCAACGTAC TTTTCTAGGA CGTTTATGA AAACCTCCCT TCGCCGAATC GGAAGAGATA TTCCCATCAC AGGTACCTAA

170 180 190 200 210 220 230 240  
 CCATAATGGA GTTTTGCTTG ACAGATGTTG TAATCAAGGG CCTACTCTAA CAGTGATTTA TAGTGAAGAT CATATTATTG  
 GGTATTACCT CAAAACGAAC TGTCTACAAC ATTAGTTCCC GGATGAGATT GTCACCTAAAT ATCACTTCTA GTATAATAAC

250 260 270 280 290 300 310 320  
 GAGCATATGC AGAAGAGGGT TACCAGGAAA GAAAGTATGC TTCCATCATC CTTTTTGAC TTCAAGAGAC TAAATTTCA  
 CTCGTATACG TCTTCTCCCA ATGGTCCTTT CTTTCATACG AAGGTAGTAG GAAAAACGTG AAGTTCTCTG ATTTTAAAGT

330 340 350 360 370 380 390 400  
 GAATGGAAAC TAGGACTATA TACACCAGAA ACACGTGTTTT GTTGTGACGT TGCATAATAT AACTCCCCAA CTAATTTCCA  
 CTTACCTTTG ATCCTGATAT ATGTGGTCTT TGTGACAAA CAACACTGCA ACGTTTATA TTGAGGGGTT GATTAAAGGT

410 420 430 440 450 460 470 480  
 GATAGATGGA AGAAATAGAA AAGTGATTAT GGACTTAAAG ACAATGGAAA ATCTTGGACT TGCTCAAAAT TGTACTATCT  
 CTATCTACCT TCTTTATCTT TTCACCTAATA CCTGAATTTT TGTACCTTT TAGAACCTGA ACGAGTTTGA ACATGATAGA

490 500 510 520 530 540 550 560  
 CTATTCAGGA TTATGAAGTT TTTTCGATGCG AAGATTCACT GGACGAAAGA AAGATAAAG GGGTCATTGA GCTCAGGAAG  
 GATAAGTCCT AATACTTCAA AAAGCTACGC TTCTAAGTGA CCTGCTTTCT TTCTATTTTC CCCAGTAACT CGAGTCCTTC

Fig. 4b

570 AGCTTACTGT CTGCCCTTGAG AACTTATGAA CCATATGGAT CCCTGGTTCA ACAAATACGA ATTCTGCTGC TGGGTCCAAT 640  
 TCGAATGACA GACGGNACTC TTGAATACCT GGTATACCTA GGGACCAAGT TGTATTGCT TAAGACGACG ACCCAGGTTA

650 TGGAGCTGGG AAGTCTAGCT TTTTCAACTC AGTGAGGTCT GTTTTCCAAG GGCATGTAAC GCATCAGGCT TTGGTGGGCA 720  
 ACCTCGACCC TTCAGATCGA AAAAGTTGAG TCACTCCAGA CAAAAGGTTTC CCGTACATTG CGTAGTCCGA AACCAACCCGT

730 CTAATACNAC TGGGATATCT GAGAAATATA GGACATACTC TATTAGAGAC GGGAAAGATG GCAATACCT GCCATTTATT 800  
 GATTATGTTG ACCCTATAGA CTCTTCATAT CCTGTATGAG ATAATCTCTG CCTTTCTAC CGTTTATGGA CCGTAAATAA

810 CTGTGTGACT CACTGGGGCT GAGTGAGAAA GAAGGCGGCC TGTGCATGA TGACATATCC TACATCTTGA ACGGTAACAT 880  
 GACACACTGA GTGACCCCGA CTCACCTCTT CTTCGCGCGG ACACGTACCT ACTGTATAGG ATGTAGAACT TGCCATTGTA

890 TCGTGATAGA TACCAGTTTA ATCCCATGGA ATCAATCAAA TTAAATCATC ATGACTACAT TGAATCCCCA TCGTGAAGG 960  
 AGCACTATCT ATGGTCAAAAT TAGGGTACCT TAGTTAGTTT AATTAGTAG TACTGATGTA ACTAAGGGGT AGCGACTTCC

970 ACAGAATTCA TTGTGTGGCA TTTGTATTGG ATGCCAGCTC TATTGAATAC TTCTCCTCTC AGATGATAGT AAAGATCAAA 1040  
 TGTCTTAAAGT AACACACCGT AACATATAAC TACGGTCCGAG ATAACTTATG AAGAGGAGAG TCTACTATCA TTTCTAGTTT

1050 AGAATTGAA GGGAGTTGGT AAACGCTGGT GTGGTACATG TGGCTTTGCT CACTCATGTG GATAGCATGG ATCTGATTAC 1120  
 TCTTAAGCTT CCTCAACCA TTGCGGACCA CACCATGTAC ACCGAAACGA GTGAGTACAC CTATCGTACC TAGACTAATG

Fig. 4c

1130	1140	1150	1160	1170	1180	1190	1200
AAAAGGTGAC	CTTATAGAA	TAGAGAGATG	TGTGCCTGTG	AGGTCCAAGC	TAGAGGAAGT	CCAAAGAAAA	CTTGGATTG
TTTTCACATG	GAATATCTTT	ATCTCTCTAC	ACACGGACAC	TCCAGGTTG	ATCTCCTTCA	GGTTTCTTTT	GAACCTAAAC
1210	1220	1230	1240	1250	1260	1270	1280
CTCTTTCTGA	CATCTCGGTG	GTTAGCAATT	ATTCTCTGTA	GTGGGAGCTG	GACCCCTGTAA	AGGATGTTCT	AATTCCTTCT
GAGAAAGACT	GTAGAGCCAC	CAATCGTTAA	TAAAGGAGCT	CACCCCTCGAC	CTGGGACATT	TCCTACAAGA	TTAAGAAAAA
1290	1300	1310	1320	1330	1340	1350	1360
GCTCTGAGAC	GAATGCTATG	GGCTGCAGAT	GACTTCTTAG	AGGATTGCC	TTTTGAGCAA	ATAGGGAATC	TAAGGGGAGGA
CGAGACTCTG	CTTACGATAC	CCGACGTCTA	CTGAAGAATC	TCCTAAACGG	AAAACCTCGTT	TATCCCTTAG	ATTCCCTCCT
1370	1380	1390	1400	1410	1420	1430	1440
AATTATCAAC	TGTGCACAAG	GAAAAAATA	GATATGTGAA	AGGTTCAAGT	AAATTCCTC	ACATCACAGA	AGATTAAAAAT
TTAATAGTTG	ACACGTGTTT	CTTTTTTTTAT	CTATACACTT	TCCAAGTGCA	TTTAAAGGAG	TGTAGTGTCT	TCTAATTTTA
1450	1460	1470	1480	1490	1500	1510	1520
TCAGAAAGGA	GAAACACACAG	ACCAAAGAGA	AGTAACATAAG	ACCAAAGGGA	TGTGTTTTAT	TAATGTCTAG	GATGAAGAAA
AGTCTTTTCT	CTTTTGTGTC	TGGTTTCTCT	TCATTGATTC	TGGTTTCCCT	ACACAAAAATA	ATTACAGATC	CTACTTCTTT
1530	1540	1550	1560	1570	1580	1590	1600
TGCATAGAAC	ATTGTAGTAC	TTGTAAATAA	CTAGAAATAA	CATGATTTAG	TCATAAATTG	GAAAAATAAT	AATAATTTTT
ACGTATCTTG	TAACATCATG	AACATTTATT	GATCTTTATT	GTACTAAATC	AGTATTAACA	CTTTTATATA	TTATTAAAAA
1610	1620	1630	1640	1650	1660	1670	1680
CTTGGATTTA	TGTTCTGTAT	CTGTGAAAAA	ATAAATTCT	TATAAAAAAA	AAAAAAAAAA	AAAAAAA 3'	TTTTTTT 5'
GAACCTAAT	ACAAGACATA	GACACTTTTT	TATTAAAGA	ATATTTTTTT	TTTTTTTTTT	TTTTTTTTT	TTTTTTTTT

Fig. 5

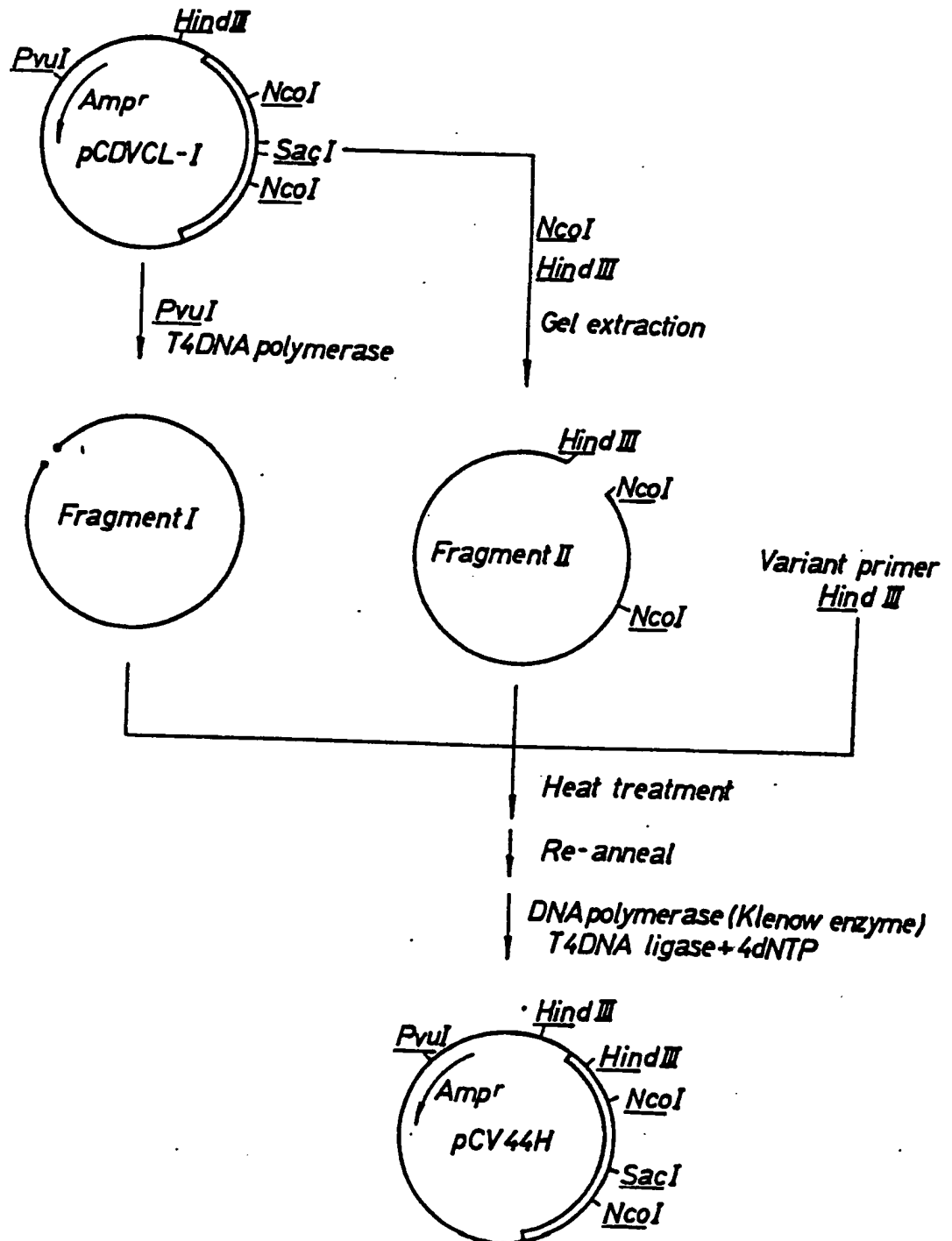


Fig. 6

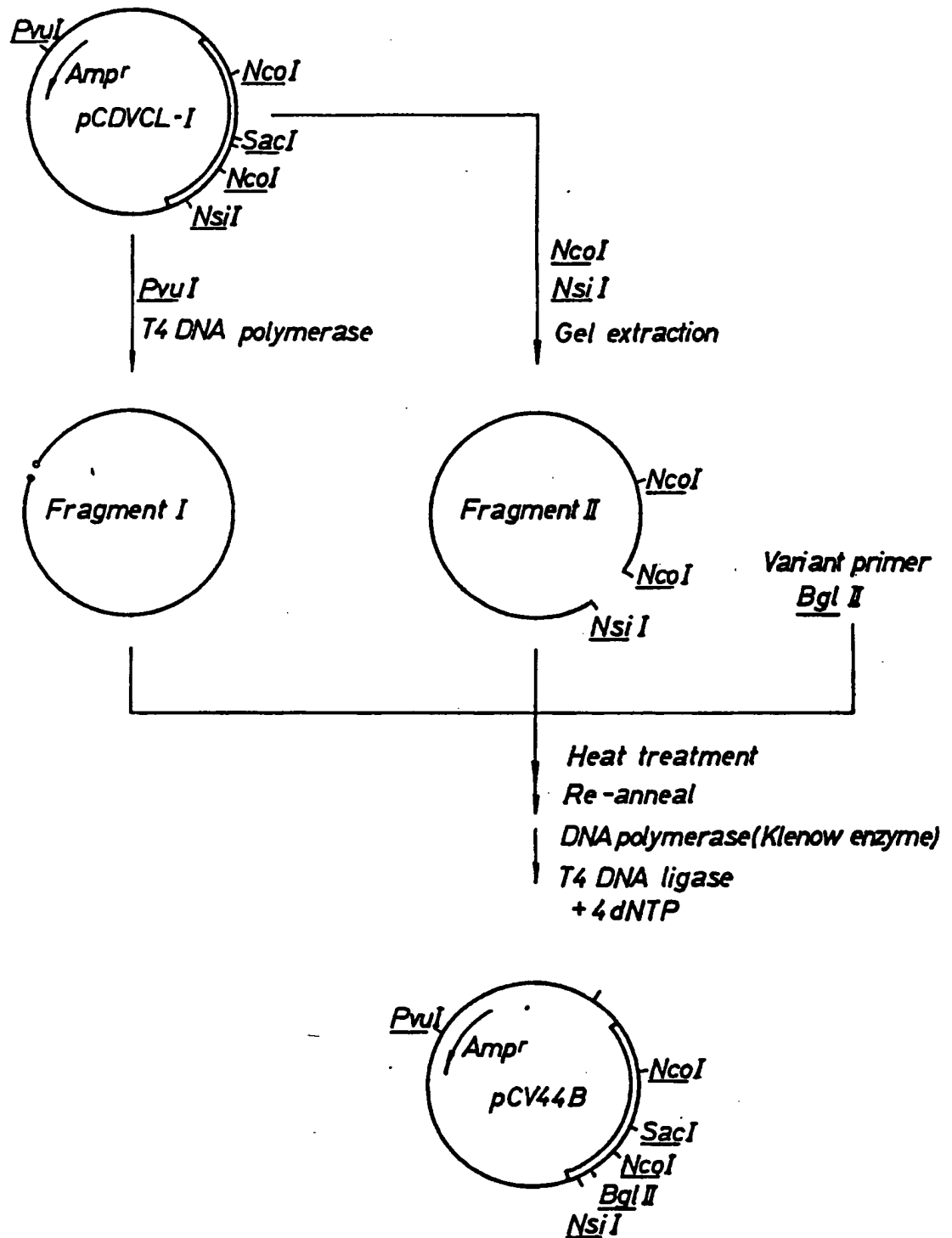


Fig.7

